भाकृअप
ICAR

CIFA

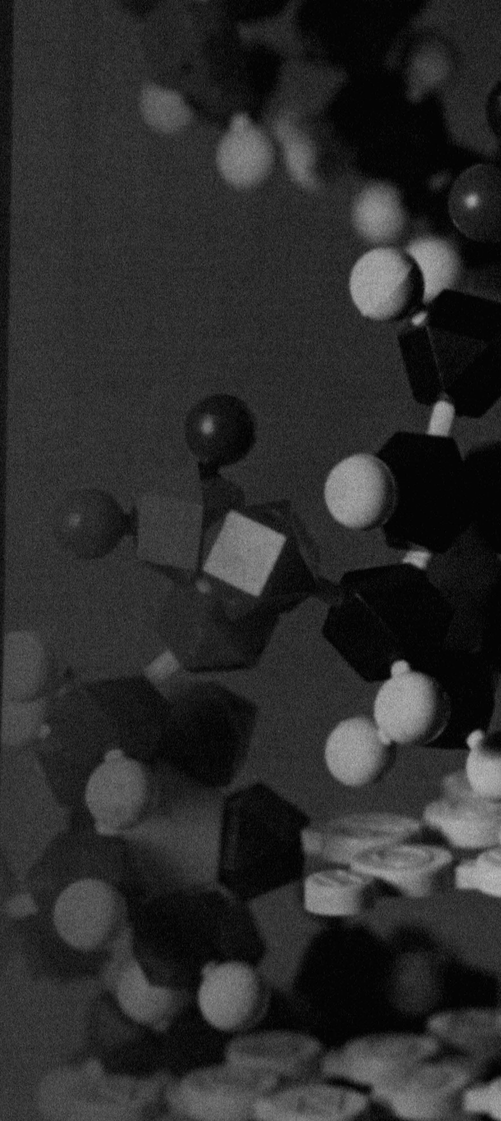
National
Workshop-cum-Training
on

BIOINFORMATICS AND INFORMATION MANAGEMENT IN AQUACULTURE

19-21 March, 2009

Nagesh Kumar Barik
Co-ordinator
Organising Secretary

Ambekar E. Eknath
Director



Sponsored by
Department of Biotechnology
Ministry of Science & Technology, Govt. of India, New Delhi

Organised By
Bioinformatics Centre on Aquaculture
Central Institute of Freshwater Aquaculture
(Indian Council of Agricultural Research)
Kausalyaganga, Bhubaneswar, Orissa

भा कृ अ प
ICAR

Bioinformatics and Information Management in Aquaculture

National Workshop cum Training

19-21 March, 2009

Edited by
Nagesh Kumar Barik

Assistance in compilation
D. P. Rath
Prasanti Mishra
Manas Kumar Behera
Krushna Chandra Parida
Srikanta Samal

Sponsored by

Department of Biotechnology
Ministry of Science and technology
Government of India, New Delhi

Organised by

Bioinformatics centre on Aquaculture
Central Institute of Freshwater Aquaculture
Indian Council of Agricultural Research
Kausalyaganga, Bhubaneswar

PROFILE OF CENTRE

BIOINFORMATICS CENTRE ON AQUACULTURE
Sub-Distribution Centre of Biotechnology Information Network
Department of Biotechnology
Ministry of Science and Technology
Government of India

PROJECT TEAM

Nagesh Kumar Barik
Scientist (Sr. Scale)
Coordinator

G. S. Saha, PhD
(Senior Scientist)

D. P. Rath, T-5

Nirupama Panda,T-6

CENTRAL INSTITUTE OF FRESHWATER AQUACULTURE
Kausalyaganga Bhubaneswar

CONTENTS

Ambekar E. Eknath, PhD
Director
Central Institute of Freshwater Aquaculture
Bhubaneswar, Orissa, India

FOREWORD

The central Institute of Freshwater Aquaculture, Bhubaneswar has been pivotal in developing and disseminating aquaculture technology since last four decades. These technologies are developed from back up support provided by the high quality research in the field of biotechnology, bioinformatics and other frontier research. The institute is a leading centre in the country in terms of the research in the field of the bioinformatics as evidenced from the fact that the Bioinformatics Centre of CIFA is only such centre in the fisheries sector in the country. The centre is uniquely contributing to the sector by providing infrastructure, information and data base support to the students and scientists. The researchers in the field are immensely benefited by the ongoing workshop and training activities of the centre.

Since 1990 the centre has already conducted 11 National Workshop-Cum-Training Programmes to update the researchers on the latest developments in the field of the biotechnology, bioinformatics, statistic and information technology. These programmes also provide a platform for the interaction of the researchers and students working in the related field.

Continuing with our traditions, this year also we are organizing the "National Workshop-Cum-Training Programme on Bioinformatics and Information Management in Aquaculture" during 19-21 March, 2009. The lectures and presentations of these workshops are compiled in this manual to make a ready reference study materials for the participants. I hope this manual will be highly useful source of information.

Ambekar E. Eknath

Nagesh Kumar Barik
Scientist
Coordinator
Bioinformatics Centre on Aquaculture
Central Institute of Freshwater Aquaculture
Bhubaneswar

ACKNOWLEDGEMENT

Bioinformatics centre of CIFA is highly grateful to the Department of Biotechnology, Ministry of Science and Technology, Government of India for supporting in creating and developing the centre since 1990. I am extremely thankful to Dr. T. Madan Mohan, Adviser, DBT for continuing his support and encouragement to carry out the mandated work of the centre.

We are indebted to all the resource persons from CIFA, Bhubaneswar; OUAT, Bhubaneswar; IARI, IASRI, New Delhi; CIFRI, Barrackpore; NIT, Rourkela; KIIT, Bhubaneswar; Berhampur University, IIMT (CSIR), Bhubaneswar; and many other not attached to any institute in investing there time and mind to generate knowledge resource useful to the students and teachers working in the field of bioinformatics. I am particularly grateful for sending the lecture notes in time to enable me to the prepare the manual for the participants to carry it with them.

Centre acknowledges all the supports and encouragement by Dr. Amber E. Eknath, Director CIFA for organsing the workshop and preparing manual in time.

Nagesh Kumar Barik

BIOINFORMATICS CENTRE FOR AQUACULTURE-BUILDING OVER STRENGTH

Nagesh K. Barik[1]

1. Summary of the progress made by the centre from its inception

The Bioinformatics Centre on Aquaculture established during 1990-91 at Central Institute of Freshwater Aquaculture (ICAR), Kausalyaganga, Bhubaneswar is a Distributed Information Sub-Centre (Sub-DIC) under Biotechnology Information System (BTIS) Network of the Department of Biotechnology, Government of India. The centre is specialized in the field of aquaculture and serves as an information source in the country. The centre functions as an information base in the field of aquaculture in the country. The scientists working in the field get ready access to the computer based information on resources, databases. Besides these the centre is also providing computer based storage and retrieval system. It also provides communication link and is also developing software packages. It is regularly conducting training programme to develop manpower and to create awareness about computerized storage and retrieval system among researchers, planners, and developers working in the field of aquaculture.

1. Major research activities in Biotechnology and its related fields in your institution

- Age effect of parents, rate of inbreeding, studies on vitellogenin and serum immunoglobulins in carps.
- Purification and characterization of serum immunoglobulins from catla (*Catla catla*) mrigal (*Cirrhinus mrigala*) and some medium carps.
- Sustainable genetic improvement of Rohu (*Lebeo rohita*) for growth through selective breeding.
- DNA fingerprinting of fish microbial pathogens using RAPD-PCR.
- Development of DNA markers in Indian major carp, *Labeo rohita* for studying population genetic variation and identification of trait associated genes
- Cryopreservation and utilization of male gametes of improved rohu (Jayanti).
- Investigation on genotoxic effects of various pesticides, insecticides and weedicides used in agriculture and aquaculture of Indian major carp *Labeo rohita*.

2. Area of Specialization of the centre : **Aquaculture**

3. Major activities of the centre during the year 2004-08

[1] **Correspondence:** Nagesh Kumar Barik, Scientist, Coordinator, Bioinformatics Centrre, CIFA, Bhubaneswar; nagesh.barik@cifamail.in

Creation of Electronic Publication

The following CD ROMs were designed, developed and created at the Centre.

(1) **National workshops cum training on Bioinformatics and Statistics in Aquaculture Research** (Proc. of workshop 99 & 2K which contains 46 articles from various professionals across the country covering Statistics and IT application in Aquaculture research.

(2) **Cifa at a Glance, 2001**. The CD contain all round activities of CIFA in research, training & extension in the field of freshwater aquaculture development in the country. There are about 600 self explanatory photographs depicting infrastructure, cultivable species, breeding, culture, harvest, marketing, training, extension, etc.

(3) **Webpage of CIFA, 2005**. The CD ROM of webpage contains the information about the Centre like infrastructure, administrative structure, divisions, projects, scientific technical manpower, products, etc.

(4) **Fish Disease Database Information System, 2000.** CD was made in collaboration with IVLP- TAR-NATP- of CIFA. The CD-ROM being the first of its kind contains Disease Database Information System for aquaculture farms. It includes standardized definitions of disease events, systematic classification of diseases, standard disease indices for recording and processing of epidemiological data. The CD-ROM contains 400 pages, which includes 250 photographs depicting fish disease events.

(5) **National workshops cum training on Bioinformatics and Statistics in Aquaculture Research** (Proc. of workshop 2000 & 2001 which contains 31 articles from various professionals across the country covering Statistics and IT application in Aquaculture research.)

(6) **National workshops cum training on Bioinformatics and Statistics in Aquaculture Research** (Proc. of workshop 2002 which contains 23 articles from various professionals across the country covering Statistics and IT application in Aquaculture research.)

(7) **National workshops cum training on Bioinformatics and Statistics in Aquaculture Research** (Proc. of workshop 2003 which contains 30 articles from various professionals across the country covering Statistics and IT application in Aquaculture research.)

(8) **CIFA, Home of Aquaplosion, 2003** – The activities of CIFA

(9) **National workshops cum training on Recent Developments in Bioinformatics and Statistics in Aquaculture** (Proc. of workshop 2004 which contains 29 articles from various professionals across the country covering Statistics and IT application in Aquaculture research.)

(10) **Book on Bioinformatics and Statistics in Fisheries Research, 2004.** The 460 page book contains 66 articles from various professionals across the country covering Statistics and IT application in Aquaculture research

4. Manual publication

Barik Nagesh.(2009) Proc. National Workshop cum training programme on Bioinformatics and information management in Aquaculture held at CIFA during 19-21 March, 2009, Kausalyaganga, Bhubaneswar

Roy, A. K (2006) Proc. National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2006 February, 7-10

Antony .A, D.P. Rath and P.K. Satapathy, (2004) Proc. National Workshop cum Training Programme on Recent Developments in Bioinformatics and Statistics in Aquaculture" held at CIFA, Kausalyaganga during -10-13, Feb., 2004. pp.270.

Antony .A, D.P. Rath and P.K. Satapathy, (2003) Proc. National Workshop cum Training Programme on Application of Bioinformatics and Statistics in Aquaculture" held at CIFA, Kausalyaganga during 04-07, Feb., 2003. pp.270.

Antony.A, D.P. Rath and P.K. Satapathy (2002 Proc. National Workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research held at CIFA, Kausalyaganga during 08-11, Jan., 2002. pp..185.

Roy, A.K; A. Antony, D.P. Rath and P.K. Satapathy (2001). Proc. National Workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research held at CIFA, Kausalyaganga during 16-19, Jan., 2001. pp..258.

Roy, A.K; D.P. Rath and P.K. Satapathy (2000). Proc. National Workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research held at CIFA, Kausalyaganga during 8-11, Feb., 2000. pp..200.

Roy, A. K., P. K. Satapathy, D. P. Rath, Ramesh Dash (1999). Proc. of 'National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research' held during February 2-5, 1999 at CIFA, Bhubaneswar. pp.175.

Rout M., A. K. Roy, N. K. Acharya, P. K. Satapathy (1998). Proc. National Workshop on Information Technology in Aquaculture Research held during February, 10-13, 1998 at CIFA, Bhubaneswar. pp.88.

Rout, M., A. K. Roy, N. K. Acharya, P. K. Satapathy (1997). Proc. Networking and Biological Data Analysis held during February, 4-6, 1997 at CIFA, Bhubaneswar. pp.137

Rout, M, N. K. Acharya, Senapati, P. K. Satapathy (1994). Proc. Perspectives in Bioinformatics and Its Application to Aquaculture held during 22-26 February, 1994 at CIFA, Kausalyaganga, Bhubaneswar. pp.132.

5. Infrastructure facilities in the Bioinformatics Centre

a. Computer & Communication facility

Hardware	Softwares
• IBM Server	• MS Office 97, 2000
• PC's	• Novel Netware 4.1
• Multimedia PC	• SPAR1
• Server	• SAS

- CD Mirror server for CDROM databases
- Macintosh SE
- 3 KVA UPS
- 1 KVA UPS
- Dot Matrix Printers
- HP DeskJet
- HP LaserJet
- Cannon Bubble jet Printers
- LCD
- Color Scanner
- Modem
- CD Writer
- 5.5 KVA Generator
- FoxPro
- QPRO
- Oracle
- M.S. Project
- Visual Studio

Operating System

- DOS
- Windows 95, 98, XP
- UNIX
- Linux

Communication and Network Linkage

The centre has acquired the following things:

- BSNL Lease Line 128 kbps
- VSAT 512 kbps
- Dial up connectivity from NIC

Local Area Network (LAN)

Forty-five Computers are distributed to different divisions/sections of the Institute and connected with the BTIS centre through LAN.

b. Scientific Software packages

Quantity one (gel analysis); DNA Star; Reference Manager
UVI – Software (gel analysis); SAS; SPSS, SYSTAT

c. Databases & other information resources

Databases Procured

1. Aquatic Sciences and Fisheries Abstracts (ASFA), (from 1978- December, 2003) by Cambridge Scientific Abstracts, and Silver Platter International, NY.,
2. Fish Base, 1997. R. Frose and D. Pauly, ICLARM, Manila, European Commission and FAO, Rome.
3. The FAO Fisheries Atlas – Preliminary Release – FAO, Rome, Italy
4. Diagnosis of Shrimp Disease – V. Alday de Graindorge and T. W. Flegel – FAO, Rome and NACA, Thailand.
5. Maxims, Ecopath and Fish growth parameters software packages in diskettes, ICLARM, Manila.
6. FishStat Plus (v.2.21) – FAO, 1999.
7. FAO - ICLARM (Stock assessment tool)
8. FishstatPlus (V 2.3) – FAO, 2000

9. Archives of the Bay of Bengal Programme 1979-2000.
10 Fish Base, 2000. R. Frose and D. Pauly, ICLARM, Manila, European Commission and FAO, Rome

6. Types of services provided

The services offered include

- e-mail uploading/downloading
- internet browsing
- online and offline bibliographic databases search
- statistical data analysis
- creation of presentations
- Photo scanning
- banner printing
- laser printing
- virus scanning
- loading of software packages
- installation of computers and printers, and trouble shooting.

7. Educational activities with details of placement

- M.F.Sc. (Masters of fisheries science), Ph.D students of Orissa University of Agricultural and Technology are being regularly trained on the use of Application of Computers in Aquaculture Research. Research workers of various national reputed organizations *viz.,* Utkal University, ICMR, RRL, Regional College of Education also avail the training facility.
- Several training programme on awareness of use of computers in aquaculture were also conducted for staff members of CIFA and officials of State Fisheries, different colleges and universities of Orissa.
- Regularly students are trained in Bioinformatics, offering studentship and traineeship under BTIS project.

8. Databases Created

Databases related to aquaculture activities covering statistics, resources, bibliography, nutrition, pathology, meteorology, bio-data of Scientists and other activities related to aquaculture were created.

Database system on socio-economic status of fish farmers

The database is to provide in-depth details of socio-economic conditions of fish farmers, which will support in boosting up the scope of result oriented research and development of fish farmers. It would support the researchers and the policy makers in decision-making.

Aqua feed for pond eco system

A database on fish nutrition with its nutrient value was being developed. It maintains details about all ingredients, its nutritional value, nutritional requirements of each species for different stages and their digestive value.

Commercial Hatching of Freshwater Prawn

A database was created for storing the general information about the mother prawn and daily proceedings of a prawn hatchery in batch wise format.

9. Papers published or presented in the conferences with reference Electronic Publication (CD ROM)

The following CD ROMs were designed, developed and created at the Centre.

(1) National workshops cum training on Bioinformatics and Statistics in Aquaculture Research (Proc. of workshop 99 & 2K).

(2) CIFA - at a Glance.

(3) WebPages of CIFA.

(4) Fish Disease Database Information System.

(5) Proceeding of the National workshop cum training on Bioinformatics and Statistics in Aquaculture Research, 2001, 2002, 2003

(6) CIFA – A Home of Aquaplosion.

10. Initiation of R&D activities in Bioinformatics

- Multiple regression analysis for biophysical model development.
- Development of software and databases in fisheries.
- Survey of different software developed in the field of fisheries/aquaculture.

11. Details of Training/ Workshop organized and total no. of users trained

Sl No	Name of the Training/Workshop	Duration	No. of participant s
1	National Workshop on Perspectives in Bioinformatics and Its Application to Aquaculture	February 22-26, 1994.	11
2	National Workshop on Networking and Biological Data Analysis	February 4-6, 1997	23
3	National Workshop on Information Technology in Aquaculture Research	February 10-13, 1998	49
4	National workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research	February 2-5, 1999.	28
5	National workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research	February 8-11, 2000	33

6	National workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research	January 16-19, 2001	43
7	National workshop cum Training Programme on Bioinformatics and Statistics in Aquaculture Research	January 08-11, 2002.	34
8	National workshop cum Training Programme on Application of Bioinformatics and Statistics in Aquaculture	February 4 - 7, 2003	51
9	National Training Programme on Application of Bioinformatics and Statistics in Aquaculture	June 03 - 06,2003	15
10	National workshop cum Training Programme on Recent Development in Bioinformatics and Statistics in Aquaculture,	February 10-13, 2004	52
11	National workshop cum Training Programme on Recent Development in Bioinformatics and Statistics in Aquaculture	February 7-10, 2006	35
12	National workshop cum Training Programme on Informatics and statistics in Aquaculture research	March 24-26, 2008	45
13	National workshop cum Training Programme on Recent Development in Bioinformatics and information management in Aquaculture	March, 19-21 , 2009	50

12. Details of the study materials generated by Bioinformatics Centre

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2006 February, 7-10

Madan Mohan .T (2006,): Growth of Biotechnology Information System Network(Btisnet) in India p.p-1-6

Roy.A.K,Sarangi.N (2006, February, 7-10): Status of Bioinformatics Centre in Aquaculture,p.p-7-21

Jagdev.P.N (2006, February, 7-10): An insight into Bioinformatics,p.p-22-23

Verma.H.C (2006, February, 7-10): Artificial Neural Networks :Concepts and Applications,p.p-24-32

De.K.Rajat (2006, February, 7-10): Artificial Neural Networks and Their Applications to Reverse Engineer Gene Regulatory Networks p.p-33-43

Kumar Sunil (2006, February, 7-10): Solvent Accessibility of proteins,p.p-44-50

Rao.A.R (2006, February, 7-10): Hidden Markov Models for Biological sequences, p.p-51-61

Seth Sohan,Jana Arun (2006, February, 7-10): Fish Freshness Determination Using Electronic Nose p.p-62-72.

Thainese.J.Fr,S.j and Kannan.V (2006, February, 7-10): Design and Techniques of Data Mining in Biotech Research,p.p-73-89.

Rai Anil (2006, February, 7-10): Data Warehouse and its applications in Agriculture p.p-90-99.

Pathak.A.K and Chaturvedi Reeta (2006, February, 7-10): Computer Networking :Basics and Essentials,p.p-100-121.

Balakrishnan.M (2006, February, 7-10): Recent Trends in Networking Concepts and Communications Technologies p.p-122-129.

Antony Acushla (2006, February, 7-10): Software Metrics: a tool to Achieve Agricultural Software Qualityp.p-130-145.

Pillai.R.Bindu(2006, February, 7-10): Software Metrics: a tool to Achieve Agricultural Software Quality p.p-146-157.

Swain.K.Saroj,Roy.A.K,Sarangi.N (2006, February, 7-10): Information on Freshwater Ornamental Fish Farming-Its Scope for Upliftment of Rural Livelihood,p.p-158-179.

Swain.P, S.Dash and.Mishra. B.K (2006, February, 7-10): Fish and Shellfish Diseases in India , p.p-180-191.

Meher.P.K, Barat.A & Das.P (2006, February, 7-10): Popgene Software for Population Genetic Analysis ,p.p-192-197.

Subudhi .N. Rabi (2006, February, 7-10): Knowledge management and Role of Information Technology , p.p-198-2002.

Sankar Muruthi G.R, Mishra P.K, Osman.Md, Vittal.K.P.R, Ravinra Chary.G and Ramakrishna Y.S. (2006, February, 7-10): p.p-2003-219.

Pathak.A.K & Chaturvedi Reeta (2006, February, 7-10): Knowledge Discovery in database and Data Mining p.p-220-230.

Routray.P and Verma.D.K(2006, February, 7-10): Potential of Germplasm Conservation for Fishes through Cryopreservation and their database managementp.p-231-236.

De.H.K & G.S.Saha (2006, February, 7-10): Innovations in fisheries Extension-application of ICT p.p-237-242.

Sankar Maruthi.G.R and Girija .A. (2006, February, 7-10): Statistical procedures for Analysis, Testing and Interpretation of Aquaculture Data p.p-250-260.

Subudhi.N.Rabi. (2006, February, 7-10): Analysis of Nominal and Ordinal Data , p.p-261-263.

Mohapatra.J.K.Pratap,Prusty.k.santosh, Mukherjee.K.C(2006, February, 7-10): Modeling of Multiobjective Freshwater Aquaculture, pp-264-275.

SaxenaR Ravi and Roy.A.K. (2006, February, 7-10): Experiments in Aquaculture: Their Logical Design and Interpretation Using analysis of Variance p.p-276-281.

Panda Kumar Arun(2006, February, 7-10): Path Analysis –A Path to Better Researchp.p-282-284.

Saxena R Ravi and Roy.A.K (2006, February, 7-10): Poverty Analysis and Measurements-some basic concepts p.p-285-292.

Roy.a.K, Saha.G.S,Sarangi.N(2006, February, 7-10): Socio-Economic,Educational and Demographic Profile of fish Farmers of Kolleru Lake ,Andhra Pradesh , p.p-293-318.

Mohapatra Das Kanta(2006, February, 7-10): Statistical Programmes in Quantitative Genetics and Selective Breeding studies p.p-319-326.

Satapathy P.K, Panda.N,Jena.N and Mishra.S(2006, February, 7-10): Data Analysis Using Software Packages , p.p-327-341.

Katiha.K.Pradeep, Jena.J.K,and Chakraborty Chinmoy (2006, February, 7-10): The production and marketing of Freshwater Aquaculture Products:An Economic appraisal p.p-342-356

Mishra bibhudatta,Muduli.H.K, Jena.J.K and satapathy.P.K(2006, February, 7-10): Statistics of Aquaculture farm facilities and Possible Scope for ICT applicationp.p-357-365

Ramesh D.B. (2006, February, 7-10): Mapping of fish Research in India:A Study based on Scopus Database p.p-366-378.

Jena Nibedita and Panda Nirupama (2006, February, 7-10): Linear Programming and its Applicationsp.p-379-384.

Panda Nirupama and Jena Nibedita(2006, February, 7-10): Inventory Management in fish and Fish Products p.p-385-389.

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2004

Antony, A. (2004). Status of Bioinformatics Centre in Aquculture, pp.01-08.

Madhan Mohan, T.(2004). Growth of Biotechnology Information System Network (BTISnet) in India, pp.09-14.

Kumar, Sunil (2004). Basics in Bioinformatics,pp.15-25.

Rath, D. P.; Antony, A. and Parida, U. (2004). Basics of Computer, pp.26-30.

Mukhopadhyay, Abhishek; Mazumdar, Pooja Anjali and Das, Amit Kumar (2004). Protein - Protein Interactions: Analysis of the Interfacing Surface pp.31-34.

Das, P. (2004). Data Generation Resources for Bioinformatics. pp.35-37.

Parida, Upendranath (2004). Computer Networks and Internet pp.38-54.

Verma, H. C. (2004). Major issues of Computer Network Security and Privacy pp.55-65.

Singh, Karan Veer (2004). Biological Databases on the Web, pp.66-69.

Maruti Sankar, G. R.; Girija, A. and Chandrasekar, P. (2004). ,pp.70-77.

Subudhi, Rabi N.(2004). Information Systems, D.S.S. and its Life Cycle ,pp.78-90.

Fr. Thainese, S. J. (2004). From Hi-Tech to Hi Touch in the Process of Data Modeling, pp.91-102.

Rai, Anil (2004). Data Warehousing - An Introduction. pp. 103-109.

Routray, Padmanav (2004). Establishment of Germplasm Repositories for Fishes and their Database Management. pp110-112.

Mohanty, S. (2004). DNA Computing: A Step to Biochips pp.113-117.

Barat A. and Meher, P. K. (2004). Application of Computer in Molecular Genetics Analysis pp.118-119.

Antony, A. and Saradhi, K. P. (2004). Digitization of information under Integrated national Agricultural Resource Information System pp.120-129.

Antony, A. (2004). Expert Systems for Aquaculturist pp.130-136.

Sahoo, Rabi Narayan (2004). Geographic Information System: Concepts and its application in fisheries and Aqaculture pp.137-149.

Satapathy, P. K. and Antony, A. (2004). Basic Statistics Using Computers pp.150-158.

Maruthi Sankar, G. R.; and Girija, A. (2004). Statistical Procedure For Analysis, Testing and interpretation of Aquaculture Data, pp.159-167.

Subudhi, R. N. (2004). Tests of Hypotheses for Fishery Sciences and Aquaculture Research, pp.168-181.

Das, A. K. (2004). Estimation of Parameters in Two-Stage Sampling, pp.182-190.

Pal, Satyabrata (2004). Outlier Theory in Fisheries pp.191-196.

Mahapatra, Kanta Das (2004). Use of Different Statistical Programs in Quantitative Genetics and Selective Breeding Study pp.197-203.

Sarangi, Jitendriya (2004). Methods of Factorial Experiments, pp.204-210.

Saxena, Ritu R.; Saxena, Ravi R. and Shrivastava, Alok (2004). Utilization of Various Statiscal and Biometrical Techniques in Biological Classification, pp.211-225.

Panda, Dilip K. (2004). On Forecasting Aspects of Water Resources Time Series Data. pp.226-234.

Saxena, Ravi R. (2004). Techniques for Analysis of Repeated Measurements, pp.235-244.

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2003

Antony.A (2003): Status of bioinformatics Centre in aquaculture, pp-1-10.

Kumar Sunil,Parhi Snehasish and Nayak Swati (2003): Comparative Protein Structure Modelling, pp-11-17.

Das.p(2003): Data Generation Resources for Bioinformatics, 17-19.

Sandha Debi Prasad(2003): Subject sphere of Bioinformatics, pp-20-24.

Ramesh.D.B(2003). Research Trends in Aquaculture-a pilot study Based on Agricola Database,pp-25-29.

Thainese.J.Fr(2003). Data Base Design for Experimental Design and Analysis for Researchers , pp-30-41

Saradhi.K.P,Roy.A.K,Saha.G.S and Antony.A(2003). Database on Social-Economic Aspects of Crap Culture in Kolleru Lake , pp-42-48.

Misra Leepa(2003). Biodatabase-development, pp-49-58.

Mohapatra.K.Pratap,Satapathy.P.k(2003). Distributed processing and Computer Network,pp-59-66

Verma.H.C(2003). Web Technologies for Agricultural Development, pp-67-72.

Murthy lakshmi(2003). Networking Professionals Through Discussion forums on the Virtual Platform,pp-73-78.

Antony. A, Satapathy. P. K, and Rath.D.P(2003). Software for Aquaculturist,,pp-79-87.

Maji.A.K(2003). Computer simulation Modelling for Aquaculture, pp-88-99.

Antony.A(2003). Emerging Trend of Expert Systems In Aquaculture, pp-100-105.

Johnselvakumar.L,Antony.A, and A.S.Ghorpade(2003). pp-106-110

Mohanty.S(2003). Language Technology in the Services of Mankind, pp-106-110

Saxena R.Ravi, Saxena Ritu(2003). Experimental findings- What to present and How to present,pp-113-118.

Saxena R.Ravi, Saxena Ritu(2003). Basic Concepts of Regression and Correlation Technique in Characterizing associations between variables ,pp-119-132.

Panda.A.K(2003). Operations Research in Decision Making pp-133-136.

Panda.A.K(2003). An Introduction to Data Envelopment ,pp-137-139

Bhar Lalmohan(2003): Linear and Nonlinear Models pp-140-158.

Panda.k.Dillip(2003). Designing Options in Fishery Experiments, pp-164-168.

Subudhi.R.N(2003). Rudiments of Probability Distributions ,pp-169-178.

Paul.A.K(2003). Heritability of all or none Traits, pp-179-189.

Paul.A.K(2003). Heritability of all or none Traits,pp-179-189.

Das Mohapatra. Kanta(2003). Use of Different statistical Programs in Quantitative Genetics and Selective Breeding,pp-190-197

Satapathy.P.K, Antony.A(2003). Exploring Statistical Data Analysis Features in MS ,pp-198-208.

Pal.S(2003). Data Mining-Concepts-statistical,pp-209-224

Reddy.P.G(2003). Forecasting Models: Methods and Certain Applications , pp-225-237

Rao.J.K(2003). Aquaculture Information on Giant freshwater prawn with Reference to its Seed Production And Culture ,pp-238-250.

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2002

Antony, A. and Janaki Ram, K. (2002). Status of Bioinforamtics Centre in Aquacutlure, pp.1-10.

Das, Amit Kumar (2002). An Overview of Bioinformatics, pp.11-13.

Sandh, Debi Prasad (2002). How to Become A Bioinforamtics Expert, pp.14-19.

Satapathy, P. K. and Antony, A. (2002). Tips of Maintaining Personal Computers By Common Users,pp.20-24.

Ramesh, D. B. (2002). Indian Agricultural Research Output - A Study Based on AGRICOLA Database for the Years 1998 and 1999, pp.25-38.

Indira, B. C. and Manbol, Raghavendra (2002). Online and Offline (CDROM) Database Search of Literature: An Overview, pp.39-46.

Roy, A. K.; Saradhi, K. P.; Saha, G. S.; Sahoo, K. N. and Antony, A. (2002). Database on Socio-Economic Aspects of Carp Culture in Kolleru Lake, pp.47-51.

Mishra, Leepa and Antony, A. (2002). Networking-Aids To Aquaculturist, pp.52-61.

Antony, A. (2002). Aquacultural Expert System – New Trend of IT pp.63-70.

Mohapatra, Pratap K. and Antony A. (2002). Neural Network : A Future Technology for Aquaculture,pp.71-78.

Antony, A.; Roy, A. K. and Satapathy, P. K. (2002). Softwares Developed in Aquaculture - An Overview, pp.79-87.

Rath, D. P.; Saradhi, K. P. and Antony, A. (2002). Emergence of E-Commerce to Improve Aquaculture Industrial Development. pp.88-90.

Sahu, B. B.; Meher, P. K.; Mohanty, S.; Pilli, Bindu R.; Rao, K. J.; Kanaujia, D. R. and Janaki Ram, K. (2002). Evaluation of Carcass and Commercial Characteristics in Prawns, pp.91-101.

Saha, G. S.; De, H. K. and Antony, A. (2002). Management Information System - An Introduction ,pp.102-106.

Panda, Arun Kumar (2002). Analysis of Variance (Anova) For Field Research pp.107-108.

Saxena, Ravi R. (2002). Logic, Research, and Experiment - Raise Effectiveness Through Better Decision Making, pp.109-121.

Das, K. (2002). Method of Drawing Simple Random Sample and Systematic Random Sample,pp.122-126.

Pal, Satyabrata (2002). Building Up Models in Fishery Research. , pp.127-132.

Meher, P. K. (2002). Indtroduction to SAS system and its Applications in Basic Regression Diagnostic ,pp.133-137.

Bhar, Lalmohan (2002). Non-Linear Models for Forecasting Fish Production from ponds. pp.138-145.

Saxena, Ravi R. (2002). Linear and Non-linear Elationships - Use Statistical Softwares to Minize Computational Burden pp.146-174.

Das, Kumar B. and Shaoo, K. N. (2002). Role of the State in Growth of Aquaculture. pp.175-181.

Rath, S. C.; Sarkar, S.K.; Kumar, K; Antony, A. and Satapathy, P. K. (2002). Study of Carp Induced Breeding Response by Different Inducing Agents pp.182-185.

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2001

Roy. A.K and Saha.C(2001). Status of bioinformatics Centre on Aquaculture, ,pp-1-10

Hota .Ashok Kumar(2001). Tips on MS-windows, pp-11-13

Thainese.J.Fr. (2001). Revolution in Data Communication,pp-14-23

Ramesh.D.B(2001) Electronic Journals-Its Features, Issues, Myths and Realities,pp-24-30

Murthy Lakshmi(2001) Information Management in the Digital Era, pp-31-37

Sandha Debi Prasad(2001) Search Engines and its search Techniques for Internet Surfing,pp-38-49

Das.P(2001) Biotechnology Information resources on the Web,pp-50-55

Padhi.B.K,CH.Jamaki And Kundu.S.C(2001) Recent Trends in the Use of Bioingormatics in Fish Genetic Research,pp-56-60

Mohanty.S(2001) Artificial Intelligence vis-a-vismachine Intelligence, Orissa,pp-61-64

Antony.A(2001) Scope of application of Expert System in Aquaculture, pp-65-73

Panduranga Rao.J(2001) Expert Systems and simulation Studies in Agriculture/Aquaculture, pp-74-87

Antony.A,Roy.A.K,Satapathy.P.K,Rath.D.P(2001) Software in fisheries/Aquaculture, pp-88-93

Saradhi.K.P,Roy.A.K,Sahoo.K.N,Saha.G.S(2001) On Socio-Economic Database management System, pp-94-97

Sahu.B.B,Pattnaik.P,Samanta.M,Mohanty.S.P,Mishra,B.K,Roy.A.K,Saha.C(2001)

Fish Disease Database Information Management, pp-98-112

Saha.c,Dey.H.K and Saha.G.S(2001)Agricultural Technology InformationCentre-An Outline,pp-113-119

Saha.C(2001). Modeling Approach in aquaculture Engineering,pp-120-122

Saxena.R.Ravi and Roy.A.K(2001). Simulation and Modelingan Overview,pp-123-140

Pal.S(2001). Rudiments of Modelling in Fisheries,pp-141-148

Roy.A.K,satapathy.P.K,Antony.A(2001) Modeling in Aquaculture,pp-149-166

Das.A.K (2001) Sampling Methodology for Modeling for estimation of population total in Aquaculture research,pp-167-173

Suriya Rao.A.V(2001).Modeling in Agriculture, pp-174-177

Gupta, S. D.; Roy, A. K.; Rath, S. C. and Satapathy, P. K. (2001). Quantitative Analysis of Induced Breeding Experiments on *Labeo rohita* ,pp-178-182.

Saxena, Ravi R. (2001). Probit Analysis - Method of Evaluating Dose Effect Experiments,pp-183-195.

Saxena, Ravi R. and Verma, Ritu (2001). Adoption and Stability Analysis in Gentic Studies - An application through Computer,pp-196-206.

Satapathy, P. K.; Roy, A. K. and Rath, D. P. (2001). Steps in Statistical Data Analysis using MS Excel,pp-207-218.

Saxena, Ravi R. and Verma Ritu (2001). Deciding what Curve to fit - Use MS Excell as a Tool,pp-219-234.

Meher, P. K. and Mishra, Shnehasis (2001). Quantitative Methods in Genetic Studies,pp-234-239.

Mishra, Shnehasis (2001). Quantitative Methods in Limnology Research, pp-240-245.

Das, K. B. and Sahoo, K. N. (2001). Fish Production Function and Environmental Sustainability,pp-246-249.

Nakhwa. D. V. and Srinivasa Rao, M. (2001). Marine fishery survey in Orissa - Financial Viability of Various Fishing Crafts, pp-250-258.

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 2000

Roy.A.K (2000). Development of Distributed Information Sub-Centre on,pp-1-8.

Hota Ahok Kumar (2000). Fundamental of computers,pp-9-15.

Ramesh.D.B(2000). Use of CD-ROM and Online Databases,pp-16-19.

Roy.A.K(2000). CD-ROM Databases, softwares and List servers related to aquaculture Research,pp-20-27.

Sandha Debi Prasad(2000) .Surfing Aquaculture Information in Internet:search Techniques, search Engines and sites,pp-28-40

Sandha Debi Prasad (2000). Designing of Bioinformatics Database Using CDS/ISIS,pp-41-53.

Mohanty.S(2000). Ant:The Model for a Six Legged Robot pp-54-57.

Mishra.B.P(2000). Soft Systems Methodology:Ideal for Strategic planning pp-58-63.

Das Lalatendu,P.K.satapathy and panda .B.K(2000). Use of GIS in Prioritisation of Watersheds on the basis of sediment yield ,pp-64-67.

Saradhi.K.P,Roy.A.K.Sahoo.K.N and Saha.G.S(2000). Development of Socio-Economic Database of Carp Cultue at Kolleru Lake of Andhra Pradess ,pp-68-71.

Qanungo .R.K,Kundu.S.C(2000). Technology of the Millennium : Biotechnology , pp-72-76.

Subudhi.R.N(2000). Fundamental Statistical Tests for Aquaculture Research , ,pp-77-94.

Saxena.R.Ravi and Roy.A.K(2000). A Note on Correlation ,Regression and Path analysis with an Application using Computer Package , pp-95-113.

Suriya Rao.A.V(2000). Cause and effect Relationship with FORTAN Solution , ,pp-114-123.

Pal Satyabrata(2000). Design of Experiments in Aquaculture,pp-124-144.

Saxena.Ravi R(2000). Factorial Experiments-An Efficient and Informative Approach ,pp-145-158.

Das.A.K. (2000). Fundamentals of Sampling Theory and its Application in Aquaculture, pp-159-172

Chakraborty.S(2000). Sampling Methodologies applied for Estimation of Species Wise Productivity under F.F.D.A. schemes in west ,pp-173-179.

Roy.A.K, Saradhi.P.K,Sahoo.K.N and Saha.G.S(2000). Development of Sampling Frame of Cultivable Water Area in Kolleru Lake of Andhra Pradesh ,pp-180-184

Satapathy.P.K, Roy.A.K(2000). Sattistical Data Analysis in MS-Excel ,pp-185-191.

Meher.P.K, Snehasish Mishra(2000). SAS System and its applications in Aquaculture ,pp-192-194

Sahoo.K.N,Mohapatra.P, roy.A.K,Saradhi.K.P, satapathy.P.K and Saha.G.S(2000)
Quantitative Methods in Aquaculture Economics Research , ,pp-195-200

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 1997

Rout.M (1997) Local Area Networks, pp1-10

Roy.A.K (1997)Fact and Figures of Internet, pp11-19

Satapathy.P.K (1997) Electronic Mail, pp20-23

M.Rout (1997) Bioinformatics, pp24-27

M.Rout (1997) Databases and Information Management, pp28-33

Bankar. C. V (1997). Application of CD-ROM to Library and information management, pp34-38

M. Rout (1997). Internationally Developed data Analysis and management software, pp39-42

Bisoi.A.K and Mishra Jibitesh(1997) Fractal Image Compression : The ultimagte Image Compression standard, pp43-50

Mohanty.S(1997) Pattern Recognition on aquatic animals using Neural Network, pp51-59

Satapathy. P.K. (1997) Application of spreadsheets in Data analysis, pp60-62

Acharya. N.K(1997) Data Analysis Through computers by Statistical & Graphic tools , pp63-65

Ayyappan. S(1997) CIFA-In the service of the Nation, pp66-69

Saha. C (1997). Industrial aquaculture by Recirculatory and flow through Systems, pp69-74

Reddy. P. V. G. K (1997) Role of Genetics in Enhancing aquaculture productivity, pp75-78

Aravindakshan. P. K (1997). Intensive Carp Polyculture ,pp78-82

National Workshop Cum Training on Bioinformatics and Statistics in Aquaculture Research 1994

Tripathi, S. D. and Rout, M. (1994). Present Status and Future Perspectives of Bioinformatics and its Application to Aquaculture, pp.1-5.

Rao, P. S. (1994). Information Technology in Bio-Sciences, pp.6-18.

Pradhan, T. (1994). Dynamics of Fish Population pp19-27.

Rout, M. (1994). Bridging the Gap Between Aquaculture and the Information Sciences pp28-34.

Nanda, R. N. and Torasia, S. (1994). Marine Bio-information System: Satellite Remote Sensing Application for Identification of Potential Fishing Zones in the Bay of Bengal Adjoining Orissa Coastpp.35-43.

Bisoi, A. K. (1994). Programming Languages for Algebraic Manipulation pp.44-46.

Mohanty, S. (1994). Knowledge Representation in Aquaculture, pp.55-62.

Rout, M. (1994). Consolidated Information from CD-Rom: A Holistic Approach. pp.47-54.

Senapati, S. S. (1994). Database Management Systems with Special Reference to CDS/ISIS, pp.63-71.

Acharya, N. K. (1994). Linear Programming - A Tool for Aquacultural Research. pp.72-78.

Rao, D. K. (1994). Data Communication. pp.79-88.

Reddy, P.V.G.K. (1994). Genetic Engineering - A Powerful Biotichnological Tool in Modern Aquaculture, pp.89-93.

Satapathy, P. K. (1994). Desk Top Publishing: A Revolutionary Technology, pp.94-98.

Senapati, S. S. (1994). Management Information System in Agriculture, pp.99-103.

Rout, M. (1994). An Inexpensive Microcomputer Based Data Record Keeping Information System for Aquaculture, pp.104-110.

Bankar, C. V. (1994). Computer-based Library System, pp.111-117.

Ayyappan, S. (1994). Biotechnological Approaches for Increasing Aquaculture Productivity, pp.118-121.

Saha, C. (1994). Modelling Approach in Aquacultural Engineering,751002, Orissa, India during February 22-26, 1994. pp.122-126.

Mahapatra, S. K. (1994). Basic Computer Concepts & Myths, pp.127-132.

AN OVERVIEW OF GENETICS AND BIOTECHNOLOGY RESEARCH IN CIFA

P. Jayasankar[1]

1. Introduction

The present industrial status of aquaculture is due mainly to the worldwide demand of fish and fisheries products and an easy access to a cheap source of protein, essential fatty acids and vitamins. The success in such an enterprise lies in monitoring many integrating factors including the genetic potential of the cultivable species that is indeed crucial for overall production performance in aquaculture. Recent advances in our knowledge on genetics and molecular biology of aquatic organisms and development of cutting edge technologies in biotechnology are expected to assist aquaculture for meeting the demand supply gap and commercialization.

Application of molecular techniques for the study of fish genomes has gained momentum. The puffer fish *(Fugu rubripes)* genome has already been sequenced and the genomes of two other species, the zebra fish (*Danio rerio*) and the Japanese medaka (*Oryzia latipes*) are in the pipeline. Many countries have demonstrated rapid growth in genetic and biotechnological sectors. Therefore, future production in aquaculture would rely, largely, on genetic improvements with strong biotechnological knowledge base. Though interest in aquaculture biotechnology has been growing globally, India is yet to realize its full potential in contrast to many developed nations. Genetically improved aquaculture species with enhanced growth rates, higher food conversion efficiency, disease resistance and product quality would certainly increase production efficiency. The application of biotechnology to aquaculture may also help to conserve wild species and genetic resources and provide unique models for biomedical research.

Genome manipulation, selective breeding, DNA marker technology, Marker Assisted Selection (MAS), transgenic technology, disease diagnostics, microbial biotechnology and Genetic diversity and conservation are the avenues for genetic improvement of aquaculture species.

2. Biotechnological efforts of CIFA

Against this background, Central Institute of Freshwater Aquaculture (CIFA), Kausalyganga, Bhubaneswar has prioritized 8 areas of biotechnological research for enhancement of fish production and improvement of genetic quality of the products. They are:

1. Breed improvement (selective breeding and biotechnology)
2. Disease diagnostics and control
3. Transgenic & value addition

---

[1] **Correspondence**: P. Jayasankar, *Principal Scientist & Head, Fish Genetics and Biotechnology Division, CIFA, Kausalayaganga, Bhubaneswar; Email:* jayasankarp@vsnl.com

4. Stem cell technology
5. Reproductive biotechnology
6. Nutritional biotechnology
7. Microbial biotechnology
8. Nanotechnology

Selective breeding of rohu (*Labeo rohita*) through combined selection approach for growth and disease resistance against aeromoniasis is an important programme of the Institute. Sixty two full sib families from 62 sires and 33 dams of selected rohu were produced. At farm level Jayanti rohu showed advantages in terms of reduction in stocking density, feed requirement, rearing period and higher yield compared to normal rohu. Resistant line of rohu (against aeromoniasis) showed 58% higher survival compared to the susceptible line. Brand name "Jayanti" has been registered.

Molecular taxonomy, generation of genetic databases, linkage map, QTL identification and MAS are the focus areas under breed improvement through biotechnogical approach. Linkage mapping of rohu and giant freshwater prawn is progressing. So far 3 linkage groups of rohu have been obtained. Molecular diagnostics for White Tail Disease (WTD) in freshwater prawn and herpes virus in koi carp have been developed. An ambitious project on development of transgenic ornamental fish is initiated. Being a non-food item, genetic engineering in ornamental fish could invite less controversy. Further, value addition is achieved. Cloning and sequencing of strong muscle-specific promoter are progressing.

Embryonic Stem cell-like colonies were developed from embryonic cells of rohu. Partial cloning and sequencing of Spermatogonial Stem cell markers have been also carried out in rohu. Construction of testis-specific cDNA library in rohu is in the final stage. Vitellogenin is important in the reproductive physiology of fish; the gene responsible for its expression was cloned and sequenced in rohu. Genes for growth hormone and gonadotropin releasing hormone were also cloned and sequenced. Spermatozoa from dead fish were cryopreserved and found viable; this research has important implications in the conservation of fish species.

Full cDNA of Polyunsaturated Fatty acid (PUFA) synthesizing enzymes has been cloned and sequenced. It has been established that source of cellulose digesting enzyme in grass carp is both indigenous and microbial. Genetic relationship of 3 species of *Bacillus* has been delineated using 16s RFLP. Lactic acid bacteria were found to be probiotic non-specific immunostimulatory in magur (*Clarias batrachus*).

Heralding research endeavour in nano technology, scaffold materials, such as collagen, chitosan, polylactic acid and polylactic glycolic acid and their suitability for *in vitro* and *in vivo* are being screened.

3. Conclusion

Fisheries sector has been playing an important role in the Indian economy. While the share of agriculture and allied activities in the GDP is constantly declining, the contribution of fisheries sector to the GDP has been increasing on a steady pace during last five decades. The share of fisheries to Agricultural GDP has increased more impressively during the same period, i.e. from 0.84% in 1950-51 to 4.70% in 2006-2007. Further, the growth rate is more pronounced during last two decades, mainly due to the increased growth of aquaculture production.

Rapid development of cutting edge technologies in modern biology has changed the pace of research in emerging branches such as genomics, proteomics as well as bioinformatics. The "Human genome project" which deals with identification of individual genes and their functional dissection has become a phenomenal success. At present, full DNA sequence of more than hundred genomes is available. This information is extremely valuable for combating disease problems in human, improvement in agriculture and animal husbandry and above all for industrial growth. Aquaculture sector should not be left out from being benefited from these frontier technologies. The future aquaculture production is likely to rely mostly on the genetic improvement programs involving modern biotechnological tools and strategies to further strengthen the sector.

IMPROVING INFORMATION FOR DESIGNING AQUACULTURE POLICY

Dinesh Marothia[1]

"....these days the best economists don't even look at secondhand data; they get them on magnetic tape and let the computer look at them. Economists have voluntarily set for themselves the limits on data collection faced by students of ancient history." **Wassily Leontief (1970).**

1. Information management in aquaculture sector: challenges and opportunities

- Aquaculture practiced since beginning of civilization
- *But*, the concept of Aquaculture management is *new*
- Challenge for researchers and policy makers-fully realize the potential contribution of aquaculture and conserving water bodies, including valuation of ecosystem
- Appropriate policies and institutional arrangements, complex interrelationships between stakeholders, decision making at different levels for promoting technologies policies and strategies- are major challenges
- All of these require support for improved qualitative and quantitative information on the biological, social and economic consequences of current and alternative policies.
- Information management assume a greater urgency for policy makers:
- Have little time to assimilate bulky data and information
- Make decision with or without information
- Fall under political pressure to make quick decisions

2. Management of information: some conceptual understanding

- Common Vocabulary Erroneously Equates Data with Information-fails to differentiate the distinctive steps in the process by which data and information are produced.
- Lack of understanding of how the analytical process relate to data collection and to the information system

2.1 The Nature of Data and Data System

- Data system represent reality by describing empirical phenomena in quantified form.

---

[1] *President Elect for the 69th Conference, Indian Society of Agricultural Economics*

- Conceptualization, operationalization, of concept(definition of empirical variables), and, measurement-three distinct steps before producing data.
- Deficiency of any of these steps –constraints quality and characteristics of data system.
- Reliability of data refers-measurement, operationalization, and, conceptual.

2.2 The Nature of Information

- Data are not information
- Information includes analysis and interpretation of data for policy decision or problem solution context
- Information system includes a data system, analytical and other capabilities necessary to interpret data, and decision making
- Information, in its technical sense, refers to systematically generated and interpreted data and is purposely designed to meet or satisfy needs of its users

2.3 Analysis as a function of Information

- Linking data and analytical systems to theoretical concepts
- An information system is the total process by which knowledge is generated and brought to bear on social decision-public/private
- An information system establish relationship between decision maker, analytical process and design and collection of data.

2.4 The Imperative of Information System Design

- As scientists we often fail to perceive and design activities as subsets of the information-which give them their meaning and significance. This is a design failure.
- All information are problem solving or purposive.
- Conceptual base of information system must be redesigned frequently to keep up with the changes in the problem being studied.

2.5 Design of Information system

- *Development and Information*
- *Economic Structure and Information*
- *Dysfunctional Behavior of the Actors*

3. Linking information base and aquaculture policy: basic assumptions

- Processes of Aquaculture Policy-Making, Planning and Management are Dynamic.

- Involve a Continuous Flow of Information Vertically (National Economic to Sectoral Levels, National to Regional Levels and Regional to Local Decision-

Making Levels) and Horizontally (Between Farms, Associations or Unions, and Between Government Departments) to Ensure that the Policies and Plans Developed at Each Level of the Planning Hierarchy Harmonize with each other.

4. Necessary conditions to formulate effective aquaculture policies

- assess and quantify impacts and implications of aquaculture policy changes and management interventions on production, supply, demands, revenues, expenditures and related resources;
- monitor and understand the response mechanisms of farmers and markets to changing socio-economic conditions; and
- determine the most appropriate and realistic objectives for each sector, and establish the optimum framework required to achieve planning or policy objectives (priorities and nature of interventions).
- All these activities require substantial amount of physical, natural, biological and socioeconomic data from wide range of sources(local to global organizations)
- Better understanding of information requirement and opportunity mapping for cooperation among different aquaculture information management agencies
- Sharing of data and information with national and international collaborative agencies/institutions.

5. Information issues

- A Variety of information are available
- Relevance for policy design differ
- High -speed -capacity processing and analytical system to process massive data
- Transmission network
- Remote sensing ,GIS
- Data Quality Control
- Decision making support system available to a wider stakeholders

6. Does growth in AIB meet needs of policy making?

- Scientific input is often undermined
- Weak link between aquaculture development and information monitoring
- Harmonized methodology for data collection be developed;
- Statistical and non statistical information with analytical explanations should be considered in order to provide relevant time-series and reliable information for efficient aquaculture management and planning; and
- The scope of the data collected should be considered in view of the changing data needs for outputs, as well as planning processes

- Many information systems were found to have been developed on the basis of the capacity of the software, rather than on what the users needed, and the primary focus was to record data rather than analyze, disseminate or use the data and information compiled

7. Some Fundamental Deficiencies: AIB to link Policy Making

- Poor understanding of the purpose
- There is a lack of common definitions, classification and "normal" reference levels
- Under-utilization of data and information compiled
- Limited capacities (generation and storage) of national Aquaculture data and information programmes
- Weak regional aquaculture information networks
- Insufficient analysis;
- Ineffective packaging and communication; and
- Poor reliability/quality, relevance and timeliness

Fig.1: Role of Information in Policy-Making, Planning and Management

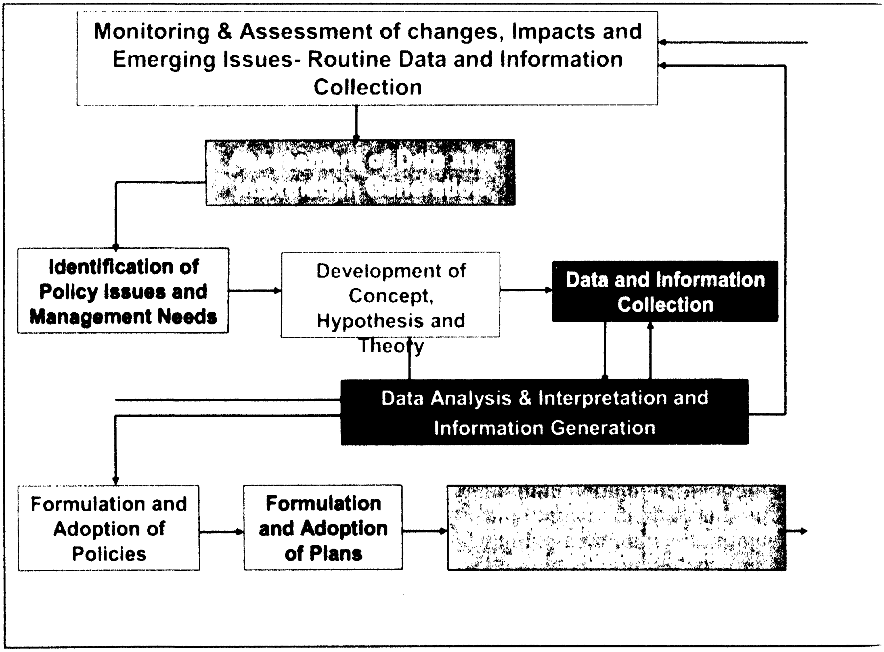
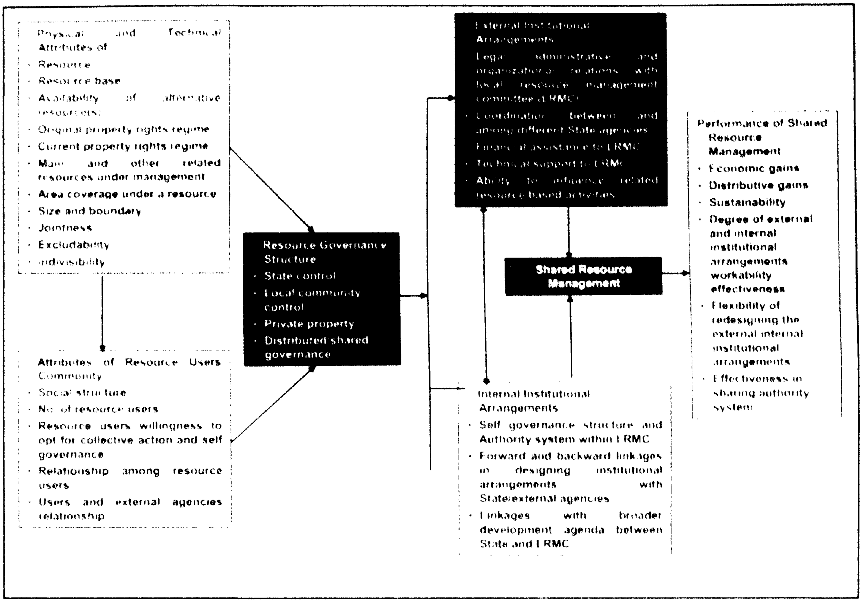


Fig.2: An Institution Framework for Analyzing Aquaculture Policy Using Information Base- An Illustration



8. Issued to be addressed to effectively link information and aquaculture policy making and planning

- Poor understanding of the purpose of data and information collection
- Under-utilization of data and information collected
- Ineffective communication and presentation
- Poor relevance, reliability, and consistency of data and information
- A lack of internationally comparable methodologies for aquaculture data and information.
- Limited capacity of national programmes
- Develop institutional framework linking information and policy governance structure-ecosystem and institutional hierarchies.
- Re-design policy attributes with continuous upgrading information-right from HH to INT Level issues.

"disdain for data collection is built in to the value and reward structure of our discipline. Ingenious efforts to tease bits of information from unsuitable data are much applauded; designing instruments for collecting more appropriate information is generally considered hack work." Rivlin (1975)

AN APPROACH TO ANALYZE SCIENTIFIC RESEARCH OUTPUT – A CASE STUDY OF ORISSA

D. B. Ramesh[1] & Sisir Mohanty

Introduction

Mapping of Indian Scientific Research output of Orissa using the online version of the Web of Science database of Thomson Scientific is presented. There were about 8082 papers published during the years 1973 to 2007 were downloaded and analysed. It can be better known from the below sample study.

Author-wise

Nayak, PL	197
Dash, AC	164
Lenka, S	162
Das, P	142
Nayak, SK	138
Kumar, A	125
Mohanty, S	125
Chakravortty, V	124
Sethunathan, N	124
Mahapatra, DP	123
Das, S	117
Khare, A	113
Rout, GR	106
Sahoo, PK	106
Dash, KC	104
Das, RP	101
Mohanty, B	100
Parida, KM	97
Phatak, SC	89
Jayannavar, AM	88
Das, D	87
Sagawa, H	85
Kumar, S	83
Nayak, TK	82

[1] Head, Library, Institute of Minerals & Materials Technology (CSIR), Bhubaneswar
Email: db_ramesh@yahoo.com

Below figure shows the collaboration with other countries

India	8796
Usa	465
Germany	285
Japan	201
Poland	183
Russia	169
Peoples R china	158
Switzerland	141
England	139
South Korea	135
Australia	123
France	123
Taiwan	101
Netherlands	93
Czech Republic	82
Brazil	76
Austria	75
Canada	75
Slovenia	68
Italy	67
Croatia	56
Sweden	49
Fed Rep Ger	15
Spain	14
Scotland	13
Bangladesh	9
Belgium	9

Document type

Article	7561
Note	853
Letter	132
Review	106
Meeting abstract	89
Editorial material	42
Correction	8
Discussion	5
Correction, addition	3
News item	2

Langauge-wise

English	8791
German	10

Year-wise

1973	53
1974	68
1975	95
1976	58
1977	130
1978	152
1979	218
1980	242
1981	277
1982	227
1983	229
1984	207
1985	149
1986	182
1987	195
1988	192
1989	188
1990	170
1991	183
1992	177
1993	190
1994	233

1995	358
1996	389
1997	377
1998	396
1999	378
2000	336
2001	360
2002	392
2003	393
2004	433
2005	365
2006	424
2007	364

City-wise

Bhubaneswar	5165
Cuttack	1640
Berhampur	1048
Sambalpur	468
Rourkela	305
Burla	179
Jaipur	108
Puri	32
Parlakhemundi	6

Institution-Wise

Utkal univ	1458
Inst phys	1282
Reg res lab	1045
Berhampur univ	847
Ravenshaw coll	743
Cent rice res inst	741

Orissa univ agr & technol	521
Sambalpur univ	332
Indian inst technol	293
Panjab univ	172
Reg plant resource ctr	154
Reg engn coll	149
Cent inst freshwater aquaculture	136
Univ frankfurt	136
Inst high energy phys	128
Bhabha atom res ctr	115
Univ sci & technol china	107
Univ tsukuba	107
Univ rajasthan	98
Orissa vet coll	93
Tata inst fundamental res	93
Univ jammu	88
Univ tokyo	88
Inst life sci	87
Osaka city univ	87
Tohoku univ	87
Seoul natl univ	86
Korea univ	85
Nagoya univ	85
Univ melbourne	85
Univ hawaii	84
Virginia polytech inst & state univ	84
Natl taiwan univ	83
Nara womens univ	82
Tokyo inst technol	82
Univ cincinnati	82
Yonsei univ	82
H niewodniczanski inst nucl phys	81
Niigata univ	81
Osaka univ	81
Sungkyunkwan univ	81
Toho univ	81
Tohoku gakuin univ	81
Chinese acad sci	80
Inst theoret & expt phys	80
Univ sydney	80

Kanagawa univ	79
Nihon dent coll	79

Source-Wise

Indian journal of agricultural sciences	270
Indian journal of chemistry section a	270
Current science	233
Indian veterinary journal	228
Journal of the indian chemical society	200
Physical review d	172
Indian journal of agronomy	156
Journal of applied polymer science	142
Physics letters b	132
Transactions of the Indian Institute of metals	122
Physical review c	121
Indian journal of animal sciences	115
Pramana-journal of physics	111
Physical review letters	99
Nuclear instruments & methods in Physics research section b	98
Physical review b	91
Hydrometallurgy	83
Physics letters a	73
Journal of macromolecular science-chemistry	71
Thermochimica acta	64
Journal of physics g-nuclear and particle physics	61
Indian journal of pure & applied mathematics	60
Journal of colloid and interface Science	60
Journal of the geological society of india	57
Journal of polymer science part a-polymer chemistry	54
Transition metal chemistry	52
Indian journal of chemistry section b	51
Indian Journal Of Marine Sciences	49
Modern Physics Letters A	48
Journal Of Physics A-Mathematical And General	47
Indian Journal Of Physics And Proceedings	44
Aquaculture	43

Indian Journal Of Pure & Applied Physics	43
Journal Of Applied Physics	43
Nuclear Physics A	43
Physical Review E	39
Biologia Plantarum	38
Pramana	37
International Journal Of Modern Physics A	36
Indian Journal Of Experimental Biology	35
International Journal Of Mineral Processing	35
Journal Of Agricultural Science	35
Journal Of Scientific & Industrial Research	34
Bulletin Of Environmental	33
Contamination And Toxicology Cytobios	33
Plant And Soil	33
Asian Journal Of Chemistry	32
Solid State Communications	32
Comparative Physiology And Ecology	31
International Journal Of Chemical Kinetics	31

Major Subjects of Research

Physics, Multidisciplinary	726
Chemistry, Multidisciplinary	674
Polymer Science	470
Physics, Nuclear	442
Chemistry, Physical	433
Physics, Particles & Fields	423
Physics, Condensed Matter	419
Plant Sciences	418
Metallurgy & Metallurgical Engineering	369
Multidisciplinary Sciences	365
Agriculture, Multidisciplinary	355
Agronomy	330
Materials Science, Multidisciplinary	285
Veterinary Sciences	279
Physics, Applied	276
Environmental Sciences	233

Astronomy & Astrophysics	226
Nuclear Science & Technology	217
Chemistry, Inorganic & Nuclear	207
Physics, Mathematical	196
Physics, Atomic, Molecular & Chemical	188
Engineering, Chemical	175
Mathematics	169
Chemistry, Analytical	167
Biochemistry & Molecular Biology	151
Engineering, Electrical & Electronic	144
Soil Science	141
Instruments & Instrumentation	129
Toxicology	129
Genetics & Heredity	128
Biotechnology & Applied Microbiology	126
Agriculture, Dairy & Animal Science	121
Fisheries	115
Marine & Freshwater Biology	110
Mining & Mineral Processing	106
Cell Biology	103
Ecology	103
Microbiology	101
Public, Environmental & Occupational Health	96
Geosciences, Multidisciplinary	93
Tropical Medicine	93
Mineralogy	91
Biology	89
Chemistry, Organic	84
Mathematics, Applied	75
Pharmacology & Pharmacy	71
Geriatrics & Gerontology	69
Engineering, Industrial	67
Water Resources	66
Immunology	65

HIGHLY CITED PAPERS

Abashian A, Abe K, Abe R, Abe K, Adachi I, Ahn BS, Ahn HS, Aihara H, Akatsu M, Akhmetshin R, Alexander JP, Alimonti G, An Q, Aoki K, Asai M, Asano Y, Aso T, Aulchenko V, Bakich AM, Banas E, Behari S, Behera PK, Beiline D, Bondar A, Bozek A, Browder TE, Casey BCK, Chang P, Chang YH, Chao Y, Chen HS, Chen

HF, Chen YQ, Cheon BG, Chidzik S, Choi S, Choi Y, Dneprovsky L, Doi Y, Dong LY, Dragic J, Drutskoy A, Eidelman S, Enomoto R, Everton CW, Fang F, Fujii H, Fujimoto K, Fujita Y, Fukunaga C, Fukushima M, Funahashi Y, Garmash A, Gordon A, Gotow K, Goriletsky VI, Grinyov BV, Guler H, Guo RS, Haba J, Haitani F, Hamasaki H, Hanada H, Hanagaki K, Hara K, Hara T, Haruyama T, Hastings NC, Hayashi K, Hayashii H, Hazumi M, Heenan EM, Higashi Y, Higuchi T, Hikita S, Hirai T, Hirano H, Hirose M, Hitomi N, Hojo T, Hoshi Y, Hoshina K, Hou WS, Hsu SC, Huang HC, Huang TJ, Ichizawa S, Igarashi S, Igarashi Y, Iijima T, Ikeda H, Ikeda H, Inami K, Inoue Y, Ishikawa A, Ishino H, Itami S, Itoh R, Iwai G, Iwai M, Iwasaki H, Iwasaki Y, Jackson D, Jang HK, Jalocha P, Jones M, Joo KK, Kagan R, Kakuno H, Kaneko J, Kang JS, Kang JH, Kani T, Kapusta P, Kasami K, Katayama N, Kawai H, Kawai H, Kawai M, Kawamura N, Kawasaki T, Kichimi H, Kim CH, Kim DW, Kim HJ, Kim H, Kim H, Kim SK, Kinoshita K, Kobayashi S, Kobayashi T, Koike S, Kondo Y, Korotushenko K, Kumar S, Kuniya T, Kurihara E, Kuzmin A, Kwon YJ, Lange J, Lee MC, Lee MH, Lee SH, Leonidopoulos C, Li J, Li Y, Liu HM, Liu T, Lu RS, Lyubinsky VR, Makida Y, Mamada H, Manabe A, Mao ZP, Marlow D, Matsuda T, Matsubara T, Matsumoto S, Matsumoto T, Matsuo H, Mindas C, Miyabayashi K, Miyake H, Miyata H, Mohapatra A, Moffitt LC, Moloney GR, Moorhead GF, Morgan N, Mori S, Murakami A, Murakami T, Nagai I, Nagamine T, Nagasaka Y, Nagashima Y, Nagayama S, Nakadaira T, Nakamura T, Nakano E, Nakao M, Nakajima M, Nakajima T, Nam JW, Narita S, Natkaniec Z, Neichi K, Nishida S, Nitoh O, Noguchi S, Nomura T, Nozaki T, Ogawa K, Ogawa S, Ohkubo R, Ohnishi Y, Ohshima Y, Okabe T, Okazaki N, Okuno S, Olsen SL, Ooba T, Ohshima T, Ostrowicz W, Ozaki H, Pakhlov P, Palka H, Panova AI, Park CS, Park CW, Park H, Peak LS, Peng JC, Peng KC, Peters M, Piilonen L, Prebys E, Rabberman R, Rodriguez JL, Romanov L, Root N, Rosen M, Rosanska M, Rybicki K, Ryuko J, Sagawa H, Sahu S, Saito M, Sakai Y, Sakamoto H, Sanda W, Sanpei M, Sasaki T, Sasao N, Satpathy A, Satapathy M, Sato N, Schrenk S, Semenov S, Senyo K, Sevior ME, Shakhova KV, Shen DZ, Shibuya H, Shimada K, Shpilinskaya LI, Shwartz B, Sidorov A, Sidorov V, Singh J, Stanic S, Stock R, Suda R, Sugi A, Sugiyama A, Suitoh S, Sumisawa K, Sumiyoshi T, Sung HF, Suzuki J, Suzuki JI, Suzuki K, Suzuki S, Suzuki SY, Swain S, Tajima H, Takahashi S, Takahashi T, Takasaki F, Takayama T, Tan N, Takita M, Tamai K, Tamura N, Tanaka J, Tanaka M, Tanaka Y, Tatomi T, Taylor GN, Teramoto Y, Tomoto M, Tomura T, Tovey SN, Trabelsi K, Tsai KL, Tsuboyama T, Tsujita Y, Tsukada K, Tsukamoto T, Tsukamoto T, Uehara S, Ueki M, Ueno K, Ujiie N, Unno Y, Uno S, Ushiroda Y, Usov Y, Varner G, Varvell KE, Vinograd EL, Wang CH, Wang CC, Wang MZ, Wang YF, Watanabe M, Watanabe Y, Wixted R, Won E, Xu ZZ, Yabsley B, Yamada Y, Yamaga M, Yamaguchi A, Yamaguchi H, Yamaki T, Yamamoto H, Yamanaka T, Yamaoka H, Yamaoka Y, Yamashita Y, Yamauchi M, Yan DS, Yanaka S, Ye SW, Yin ZW, Yokoyama M, Yokoyama T, Yoshida K, Yoshimura Y, Yu CX, Yuta H, Zaslavsky BG, Zhang CC, Zhang J, Zhang SQ, Zhang ZP, Zhao H, Zhao ZG, Zheng YH, Zheng ZP, Zhilich V, Zontar D

Utkal Univ, Bhubaneswar 751004, Orissa India

The Belle detector

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH SECTION A-ACCELERATORS SPECTROMETERS DETECTORS AND ASSOCIATED EQUIPMENT
479 (1): 117-232 FEB 21 2002
Times Cited: 318

NEED AND POTENTIAL OF GERMPLASM CONSERVATION AND THEIR DATABASE MANAGEMENT

P. Routray[1] and D. K. Verma

1. Introduction

Global diversity in domestic animals and aquatic organisms is considered to be under threat worldwide. The multifarious human activities in open waters have threatened the fish diversity all round the world. Fish worldwide are in crisis. The effective population sizes in fish hatcheries are getting smaller and smaller that causes loss to farmers worldwide. The world Resources Institute has reported that nearly 70 percent of the world's marine fish stocks are over fished or are being fished at their biological limit. Apart from fish, across all live stock species there is a shrinking pool of genetic diversity. For instance, the effective population size for all dairy cattle breeds is less than 60 animals. To preserve the gene pool of unique populations by protecting them from extinction, loss of genetic variability (inbreeding) and dilution due to interbreeding with unrelated populations (hybridization) gene banks or germplasm repository can play a vital role. The germplasm repositories in fisheries could be used for: 1) restoration of threatened species; 2) brood stock improvement through genetic modification of a targeted population, and 4) information and technology development including gene pool database management. Fish population germplasm repository can be of two types: live animal banks which are maintained at multiple locations to provide a continuous source of progeny, or tissue gene banks based on gamete, zygote or somatic tissues maintained in cryogenic condition for later retrieval. In fishes, mostly sperm cells are used as source with greater accuracy in producing results (Routray *et. al.*, 2003). Gene banks may be used for decades for restoration of a species and for genetic manipulation studies. During the prolonged and regular use of gene banks, a huge amount of information may be generated and must be stored in an appropriate form in a user friendly manner. The only way all this information can be managed effectively for gene bank managers and other users is with a well designed computerized data base.

[1] **Correspondence**: P. Routray, Senior Scientist, Central Institute of Freshwater Aquaculture;Kausalyaganga, Bhubaneswar-751002,Orissa; E mail: routray30@yahoo.co.in

2. Prerequisites for Establishing a Gene Bank

Before establishing a gene bank a thorough thinking must be there about the type of gene bank to be established. For instance if a live gene bank to be established then the following things needs to be ensured: 1) A suitable land with perennial water source; 2) high embankment of ponds or tanks to avoid inter mixing of populations; 3) the species must be acclimatized to that agro climatic conditions. There are limitations of live gene banks as these are to be maintained at multiple locations, problems of mass mortality due to disease, unforeseen calamities like heavy rain fall or flood etc. The tissue gene bank has the advantages of overcoming the above problems and easy to manage.

3. Role of Germplasm Repositories

The germplasm repositories across different countries generally aimed at addressing the characterization and conservation of animal genetic resources. The following roles are ideal for germplasm repositories:

i. Understanding the population structures and demographic changes in population size, location and number of producers raising specific breeds;
ii. Acquiring germplasm (semen, embryos, ova and DNA) for storage in the repositories;
iii. Evaluation, development and implementation of cryopreservation protocols for long term germplasm cryoconservation;
iv. Development of an information database and tracking system that links genetic preservation (ex-situ and in-situ) and population management;
v. Act as a nodal centre for easy accessibility to different regions on species restoration during disasters.

4. State of the art in cryopreservation technology

Cryopreservation allows virtually indefinite storage of biological material without deterioration over a time scale of at least several years (Mazur, 1985), but probably much longer. Many important mile stones in cryobiological research could be achieved in the second half of the previous century. Much progress resulted from empirical studies. In later years, progress was also strongly stimulated by the development of fundamental theoretical cryobiology. The principle of glass transition is well applied in different cryopreservation protocols which involve the cooling of biological materials at a range of cooling rates that are fast enough to prevent slow cooling damage but are slow enough to allow sufficient dehydration of the cells to prevent intracellular ice crystal formation. The preservation of fish semen and blastomeres are a success story for germplasm conservation through cryopreservation. More than 200 fish species with external fertilization have been tested for sperm cryopreservation. The present state of art for many species of fish seems to be adequate for the purpose of gene banking and germplasm conservation.

5. Management of Gene Bank Information

A huge amount of information is required to manage the broad range of activities that are essential for an operating gene bank such as population sampling procedures, sample identification, tissue processing, cataloging, information retrieval, quality control and sample disposition. Gene pool preservation requires that a sufficient number of individuals be collected to effectively sample the total available gene pool and to achieve the gene and genotype frequencies which are representative of the source population. The actual number of individuals necessary to achieve this goal is ambiguous and debatable, but the criteria most commonly used is the smaller of a predetermined percentage of the total population (10-100%) or an upper limit number of individuals (100, 400 or more) that sampling theory indicates is sufficient (Kincaid, 1983, Ryman,1991). Once milt is collected, a sample identification system is applied to ensure positive identification of each sample during all handling procedures (species, date, time, processing, storage, monitoring and retrieval operations). A quality control program is needed to ensure that high viability is maintained. A milt replacement program is also needed to ensure that the deteriorated milt is replaced or removed for breeding programs.

Gene bank information is primarily of two types: 1) Gene bank operations information to catalog and manage the stored tissues, and 2) gene bank user information to characterize the source population and determine the most effective applications for stored tissues. The database must include the information needed to accomplish the following functions:

1. Sample and population identification procedures.
2. Collection procedures.
3. Sample processing procedures.
4. Sample storage and retrieval operations.
5. Quality monitoring.
6. Inventory maintenance.

7. Information retrieval.
8. Milt replacement program.
9. Milt discards procedures.
10. User application programmes.

The decision to preserve a specific population gene pool requires the collection of baseline information to determine population uniqueness and the traits which define this uniqueness. Baseline information includes population name, location, range and distribution, genetic characterization, breeding history, life history, habitat preferences and any specific trait pertaining to the species. Species with no cryogenic preservation history should be studied to determine a species specific cryogenic processing and storage procedures before gene banking operations can begin.

The information regarding gene bank users should be based on the population characterization data, stored milt quality and quantity data and the effective population size represented in the preserved milt. Information on population origin, breeding history, source habitat parameters, life history traits, genetic analysis and performance characteristics in culture and natural habitats should be clearly brought out. This information would help the users to match the preserved stock with the existing ones and change the management accordingly.

6. Gene Bank Information Database

A database is a collection of data arranged in a systematic order for quick search and retrieval operation in subsets. The first step involving the designing of database is crucial and the essential requirements are: development of data structure based on the identification of specific data elements. Essential information includes data needed by the gene bank managers to facilitate decisions on management and research applications where cryopreserved milt will be used.

The initial data element or trait list is developed by answering in complete details a series of questionnaires as shown in Table-1. The questions are of two types; 1) for the gene bank operations and 2) the additional information needed to assist user groups. During the development of initial trait list, representatives of all the potential users including Government officials should be involved.

Table 1. A sample questionnaire for initial trait list development for a new gene bank information.

Type:I For Gene bank operations and management		
Sl No	**Information description**	**Example traits or characteristics**
1	Population sample identification system	Population identifier, sample identifier, ampule or straw identifier etc.
2	Population sampling plan	Population size, number of collection sites, nuber of individuals collected per site, total numbers etc.
3	Milt at field	Date spawned, sample and fish identification, capture method, spawning method,, milt holding method, milt quality measurement, etc.
4	Milt transportation system?	Cooling and freezing methods, shipping temperature, transit time, transportation mode, etc.
5	Source population characteristics	Origin, breeding history, life history, reproductive traits, behavioral traits, habitat preference etc.
6	Source population genetic characteristics	Analysis method, allele frequencies, heterozygosity, distinctive characteristics etc.
7	Source fishery characteristics	Location, water body type, water quality, food source, predator species etc.
8	Gene bank processing procedure	Milt quality, additives, diluents, dilution rate, freezing procedure, storage unit type, etc.
9	Quality control and monitoring	Milt testing schedule, test sample collection procedure, fertility rate, etc.
10	Milt replacement procedures	Program identification, date removed, replacement criteria
11	Brood improvement programme	Prog. Identification, date removed, milt quality, number progeny produced, brood stock improvement results
Type:II For Gene bank user information (addition to type-I)		
1	Source population characteristics?	Additional traits and characteristics of special interest to milt user: origin, source, habitat, life history, cultural performance etc.
2	Source population status	Threatened or endangered, effective population size, disease and health status, etc.
3	Source population genetic characteristics	Source population, allele frequencies, heterozygosity, effective population size, genetic analysis etc.
4	Milt availability	Gene bank population milt inventory, gene bank milt replacement programme, restoration and enhancement etc.

Development of the information database management should be a priority when establishing a gene bank. The gene bank information management data base is essential because it is the management system that will catalog, summarize, query and retrieve information from the data set required to establish and operate gene bank. Fisheries mangers and hatchery owners using cryopreserved milt will rely on the data base for accurate information on source population, milt quality, milt availability etc. India should ensure that the genetic diversity is preserved as the Agricultural Industry appears to be headed towards a few lines of proven performers. In India the fisheries sector has one bureau (National Bureau of Fish Genetic Resources at Lucknow) for germplasm inventory, evaluation and gene banking of freshwater fishes but for marine species there is as such no organization. At Central Institute of Freshwater Aquaculture, Bhubaneswar protocols for cryopreservation of fish semen, blastomeres and other cells of freshwater fishes has been developed (Gupta *et. al.*, 1995, Routray, 2003, Routray *et. al.*, 2003) but maintenance of germplasm of many species are not being looked after as Scientists are working on the upgradation of genetic stock by different methods. In Indian fisheries, there is no full fledged fish gene bank in operation, so it is essential to establish a gene bank to save and propagate our rich fish biodiversity in these sub-continent and other places also.

7. Miscellaneous emerging reproductive technologies

Transplantation of ovarian tissue and germ cells (e.g., primordial germ cells (PGCs) or spermatogonial stem cells (SSCs) are emerging technologies with potential for future use in conservation programs. Successful transplantation of embryonic cells has been reported in some fish species. Production of xenogenic transplantation of PGCs for production of surrogate fish has been reported.

References

Gupta, S. D., Rath, S. C., and Dasgupta, S. (1995)Fertilization efficiency of cryopreserved carp spermatozoa over four years at different time intervals after thawing. Geobios, 22, 208-211.

Kincaid, H. L. (1983) Inbreeding in fish populations used for aquaculture. Aquaculture, 33:215-227

Mazur, P. (1985) Basic concepts in freezing cells. Proceedings of the 1st International Conference on deep freezing of boar sample. Uppasala, 1985 (L.

A. Johnson and K. Larsson, eds), pp. 91-111.

Routray, P. (2003) Evaluation of cryopreserved milt of a selected carp, *Labeo calbasu* for fertilization and seed production, Ph. D. Thesis, Sambalpur University, Sambalpur, Orissa.

Routray, P., Gupta, S. D., and Behera, M. K. (2003) Cryogenics in fish hatchery technology. *Fishing Chimes*, 23:7-9.

Ryman, N. (1991) Connservation genetics considerations in fisheries management. Journal of fish Biology, Supplement A 39:211-224.

APPLICATION OF REMOTE SENSING AND GIS IN AQUACULTURE AND FISHERIES IN INDIA – AN EXPERIENCE

Sanjeev Kumar Sahu[1]

1. Introduction

Natural resource management requires rapid and accurate methods for interpreting data for development and management. For the management of any natural resource, monitoring and surveying are two main factors. Inventory of inland surface water resources used for fisheries is very important for fisheries development. In the process of inventorying of water bodies, there are various aspects, which should be considered at the same time such as the location, soil, aquatic data, physico-chemical parameters, socioeconomic, livelihood, craft and gears, species composition and ownership (like government, private), fisheries activities and fish production. Surveying and monitoring of surface water resource, associated land use / land cover, vegetation, water quality, primary producers, sediment, TDS etc. could be performed by remote sensing technique.

Remote sensing is a multi disciplinary science of deriving information about an object from measurement made at distance far from the object without coming in contact with it. Remote sensing refers to the identification of earth features by detecting the characteristics electromagnetic radiation that is reflected / emitted by earth surface. Every object reflects / scatters a portion of the electromagnetic radiation depending upon its physical properties in addition; objects also emit electromagnetic radiation depending upon their temperature and emissivity. Reflectance emittance pattern at different wavelengths for each object is different; this enables identification and discrimination of objects.

Presently, the commercial availability of remotely sensed satellite data of Indian as well as foreign satellites is easily possible. These remotely sensed data have high spatial, spectral, radiometric and temporal resolution, eases the monitoring and surveying of the wetland resources. The spatial resolution of Indian satellite varies from 2.5 m to 350 m, spectral resolution varies from 1 band to 8 bands and temporal resolution varies from 2 days to 25 days.

India is an agrarian society where 70% of the population depends on agriculture, animal husbandry and fisheries. Fishery is an important sector, its planning and development largely depends on a sound and reliable data and information on resource, production, marketing and distribution. Generating reliable information basically needs identification, conceptualization, and classification and sound scientific techniques for collection, compilation and analysis of data. Geographical Information System emerged as an efficient tool to access, organize, update and analyse the information and to utilize it in an optimal way.

A proper computer based information system is the need of day to organize such complex data in order to draw reliable conclusions which may help in drawing

[1]**Correspondence:** Sanjeev Kumar Sahu Central Inland fisheries Research Institute, Barrackpore 700 120, West Bengal; email:sksahu@cifri.ernet.in

action programmes. This multidimensional nature of information on topography, physico-chemical parameters, socio-economics and fish production data can only be handled and processed by Geographical Information System (GIS).

GIS is a generic term implying the use of computers to create and display digital maps. The attribute data which describe the various features presented in maps, may relate to physical, chemical, biological, environmental, technological, social, economic or other earths surface properties. GIS allows mapping, modeling, querying, analyzing, and displaying large quantities of such diverse data, all held together within a single database. Its power and ability to integrate quantities of information about the environment and wide repertoire of tools provided to explore the diverse data.

GIS produces and reads maps. Its major advantage is that it permits identifying spatial relationship between different map features. It can create map in different scale, projection and colours. But it is not a map-making tool. It is an primarily an analytical tool that provides new ways of looking at, linking and analyzing data by projecting tabular data into maps and integrating data from different diverse resources.

2. Principle of remote sensing

Remote sensing is not a new concept. Visual perception of human eye is a best example of remote sensing. The human eye (sensor) captures light (visible radiation) from a apple (objects), placed at a distance (remote). Different objects return different amount and kind of energy in different band (range) of the EM spectrum incident upon it. This unique property depends on the property of material (structural, chemical, and physical), surface roughness, angle of incidence, intensity and wavelength of radiant energy.

The remote sensing is a science, which includes a combination of various disciplines such as optics, spectroscopy, photography, computer, electronics, telecommunication, satellite launching etc. All these technologies are integrated to act as one complete system in itself, known as remote sensing.

There are number of stages in a Remote Sensing System, working as links in complete, and each of them is important for successful operation.

3. Stages in Remote Sensing

1. Origin of electromagnetic energy (Sun, transmitter carried by the sensor).
2. Transmission of energy from the source to the surface of earth and its interaction with the intervening atmosphere
3. Interaction of energy with the earth surface (reflection/ absorption/ transmission) or self-emission.
4. Transmission of the reflected / emitted energy to the remote sensor placed on a suitable platform.
5. Detection of energy by the sensor converting into photographic image or electrical output.
6. Transmission / recording of the sensor output.

7. Pre-Processing of the data for generation of the data product.
8. Data processing and interpretation.
9. Thus the remote sensing system consists of a sensor to collect the radiation and a platform which can be satellite, rocket, aircraft and balloon on which sensor can be mounted. The information received by the sensor is suitably manipulated and transported back to earth may be telemeter or brought back through films, magnetic tape etc. The data are reformatted and processed on the ground to produce photographs, computer compatible magnetic tape (CCT), CD ROM, DVD etc.

Remote sensing is classified into three types with respect to the wavelength regions;

1. Visible and Reflective Infrared Remote Sensing,
2. Thermal Infrared Remote Sensing and
3. Microwave Remote Sensing.

4. Spectral Signature

Spectral signature is the ratio of reflected energy to incident energy as a function of wavelength. Various materials of the earth, surface has different spectral reflectance characteristics. Spectral reflectance is responsible for the color or tone in a photographic image of an object. Trees appear green because they reflect more of the green wavelength. To obtain the necessary ground truth for the interpolation of multi-spectral imagery, the spectral characteristics of various natural objects have been extensively measured and recorded. Following figure shows a typical reflectance curves for three basic types of earth surface features, healthy vegetation, dry bare soil and clear river water.

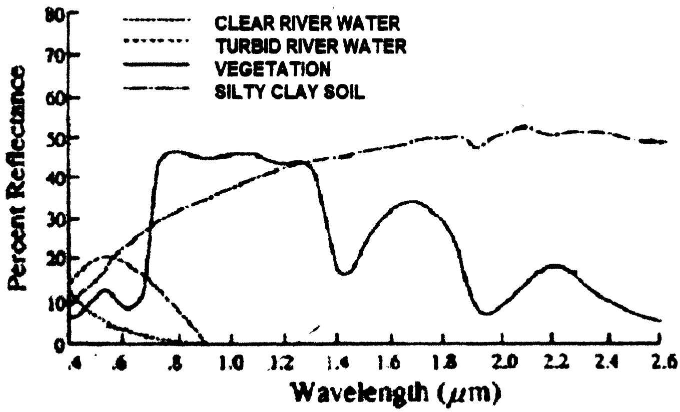


Figure 2: Spectral signature

5. Sensors

Sensor is a device that gathers energy (EMR or other), converts it into an electronic signal and presents it in a suitable form for obtaining information about the target under investigations. These may be active or passive depending on the source of energy.

Sensors used for remote sensing can be broadly classified as those operating in optical infrared (OIR) region and those operating in the microwave region. OIR and microwave sensors can further be subdivided into passive and active.

Active sensors use their own source of energy. Earth surface is illuminated through energy emitted by its own source, a part of it reflected by the surface in the direction of the sensors is received to gather the information, and passive sensors receive solar Electro-magnetic energy reflected from the surface of energy emitted by the surface itself. These sensors do not have their own source of energy and can not be used at night time, except thermal sensors. Again sensors (active or passive) could either be imaging, like camera, or sensor which acquires images of the area and non imaging types like non scanning radiometer or atmospheric sounders. There are various types of sensors, which used in different satellites.

6. Need for GIS

Development of an information system should be based on various aspects which are environmentally friendly, economically feasible and socially acceptable. For planning accurate, reliable and timely information which supports rules for management such as variety of maps, large amount of spatial data, selectively extract information relevant to a planning task is required. This type of need can be fulfilled by developing an information system based on geographical data.

7. Definitions of GIS

Geographic: Data referenced by some coordinate system to locations on the surface of the earth.

Information: Provides answers to queries about the geographic world.

System: This is the environment which allows data to be managed and questions to be posed.

A GIS is basically a computerized information system like any other database, but with an important difference that all information in this linked to a geographic (spatial) reference (latitude/longitude or other spatial coordinate system)

There are no clear-cut definitions for GIS, different people defined GIS according to capability and purposes for which it is applied. Few of them are:

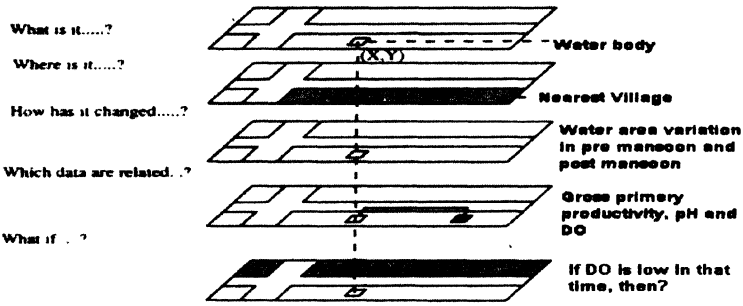
Clarke defined it as a computer system for the capture, storage, retrieval and display of spatial data, within a particular organization.

Burrough defined it as a powerful set of tools for collecting, storing, retrieving as well as transforming and displaying spatial data from the real word.

ESRI defined this as an organized collection of computer hardware, software, geographic data and personal designed to efficiently capture, store, update, manipulate, analyze and display geographically referenced information.

8. Basic questions that can be investigated using GIS

Location	-	Where is it................?
Condition	-	How is it................?
Trend	-	What has changed ....?
Routing	-	What is the best way ....?
Modelling	-	What if..........?



Required Functions of GIS

9. Acquisition of data for GIS

GIS can not work without data. The more the data more versatility and greater potential functionality of GIS. There are two type of data *viz.* primary data and secondary data. Primary data collection consists of acquiring various types of original data by a number of persons or direct methods and techniques. Secondary data represents available data or information, which may be availed in different formats.

10. Data acquisition and input

First Step - Developing database for GIS
- Acquire data
- Input to GIS

Data
- Spatial
- Non-spatial

e.g. Simple object in space - a water well location specifying any latitude/longitude (spatial). For certain application - wide range of additional information required e.g. depth, volume of water, quantity of water etc (attribute or non-spatial)

11. Mode of data Input

Primary data

Data acquired by primary collection techniques could be in many forms like photographic numeric, digitally encoded etc..

Secondary data

All secondary data represents primary data which has been converted into more accessible and processed form. The sources of secondary data at libraries, Government office, bookshops, private and public organizations etc. Secondary data which are more relevant to GIS are maps table and digital data. Mapped information is the major source of data for GIS.

Maps

Various types of maps are available. But following types of maps are important for fisheries.

- Hydrographic map
- Relief map
- Topographic map
- Political and administrative map
- District planning maps

Satellite Imageries and Aerial photographs

Visible and reflective infrared Remote Sensing
Thermal infrared Remote Sensing
Microwave Remote Sensing

Non spatial data source

It refers to numeric and textual data which has been presented in some logical and cohesive format. Important tabular data source includes census office, fisheries office, fisheries research institutes, academic research institutes, Fish marketing office, harbours and port authorities, meteorological office other than state and central government agencies and International organizations.

12. Database preparation

In any information system database is the backbone. Without proper database it is not possible to make any information system. Geographic Information System has two types of data. These two data are entirely different in nature. They are :

1. **Spatial data** (latitude/longitude for Geo-referencing, the features on a map, e.g. water body spread area, administrative districts, road), and

2. **Attribute data** (descriptive data about features, e.g. productivity of pond, population of districts, length of road etc.)

13. Data Storage and Retrieval

A GIS does not store maps. It stores data organized into database. The location data of different features (coordinates, topology) are generated during the digitization process. The attribute data of location are created separately. The GIS must provide the link between the location and attribute data. The relational database model is most suitable to ensure such linkage between the location and attribute data. The relational database model is most suitable to ensure such linkage and the database query language can be used to retrieve the data. Relational database concepts are therefore central to organizing and managing database in GIS. The specific format of data storage varies with the GIS software.

Advantages with GIS

- Consistency
- Accessibility
- Precision
- Accuracy
- Completeness
- Flexibility
- Refinability

14. Application of remote sensing and GIS in inland fisheries

14.1 Mapping of wetlands/Water body

Precise inland water resource information is vital for estimation of fish yields and production. Due to global change ad anthropogenic pressure either these resource shrinking or changing their shapes. Regular estimation of water area is very much essential to estimates the yield and production of the water resources. There are various steps involved in mapping of wetlands starting from geo referencing, base map development, image processing to tabulation. Following are the basic steps involved in water body mapping.

- Acquisition of remote sensing Image
- Geo referencing of Remote sensing Image
- Geo referencing is the process which creates and store control information
- Interpretation

14.2 Water quality

Remote sensing applications are limited to measuring those substances or conditions with influence and change optical and / or thermal characteristics of surface water properties/ water quality. Suspended sediments, chlorophylls, dissolved organic matter and temperature are water quality indicator that can change the spectral and thermal properties of surface water are most readily measured by remote sensing techniques.

Suspended solids in water produce visible change in the surface of water and in the reflected solar radiation, such changes in the spectral signal from surface water, captured by the satellite. There are significant relationship between total suspended sediment and the reflectance of Landsat TM digital data. Reservoir trophic state and Carlson trophic state index were measured in Te-Chi Reservoir in central Taiwan with the help of Landsat TM (TM1~TM4). There are significant correlation between LISS III RS digital data and total nitrogen, gross primery production and respiration.

14.3 Plankton (Chlorophyll)

Central to the aquatic food chain is phytoplankton - microscopic plants that photosynthesize chemicals in water. This process depends on the chlorophyll content of plankton. Lakes and water bodies depend mainly on their watershed for nutrients and other substances to sustain biological activities. While these nutrients and substance are required for a healthy aquatic environment, an excess of these input leads to nutrient enrichment and eutrophication of the lake. Eutrophication of a water body is usually quantified in terms of concentration of the chlorophyll contained in the algal/plankton. Plankton contains a pigment that strongly absorbs red and blue light. As plankton concentration increase, there is a corresponding rise in spectral radiances, peaking in the green. Upwelling masses of water (usually associated with thermal convection) containing phytoplankton take on green hues in contrast to the deep water with few nutrients. Remote sensing can be used to measure the chlorophyll concentration and patterns in water bodies. While measuring chlorophyll by remote sensing technique is possible, studies have also shown that the broad wavelength spectral data available on current satellites do not permit discrimination between chlorophyll and suspended sediments when suspended sediment concentrations are high due to the dominance of the spectral signal from suspended sediment. Recent research shows the relationship between chlorophyll-a and the narrow band spectral details at the red edge of the visible spectrum (Gitelson et al. 1994). Data have shown a linear relationship between chlorophyll-a and the difference between the emergent energy in the primarily algal scattering range (700-705 nm) and the primarily chlorophyll-a absorption range (675-680 nm). Laboratory and field studies using hyperspectral data have been used to develop algorithm to estimate green and blue green algae. Hyperspectral data now available from several satellite platforms, should allow better discrimination between pigments thus allowing the identification of broad algal groups.

14.4 Land use / Land cover change

Associated land use/ land cover of water body affects the water quality of the water area. Susumu ogawa and others studied water quality in Okutama lake in Japan .The purpose of the study was to examine relationship between land cover change and water quality in river. They found water quality changed by land cover change and other factors. Mangrove forest (dense, open), change in forest area and deforestation of mangrove can be assessed by remote sensing.

The long term management plan can be prepared based on the remote sensing data supported by ancillary data for habitat protection tourist attraction and eco development for the livelihood of the people. "Nal Sarovar" of Gujarat, India was monitored with the help of remote sensing for waterfowl habitat management.

14.5 Management of resource wise and location wise data

Assessment of fisheries resources and estimation of production are the most required research area. Time series data of various resources (like estuaries, reservoirs floodplain wetland) at various level (national, state and district) are being considered for this purpose. Creating this database in vector GIS has got many advantage in understanding not only temporal trends in fisheries growth but also gives spatial patterns. Further tool GIS gives will be useful in mapping these trends.

Now more than ever, decision makers at all levels need an increasing amount of information to help them understand the possible outcomes of their decision by integrating the data from different sources and types. Due to integration various statistical tools its give more liberty to decision maker and researchers to analyze the data.

14.6 Development of Inventory

In the fisheries sector of India, to develop a proper inventory is quite challengeable and Herculean task. The Indian estuaries have tremendous potential but production is quite below the predicted one. Due to lack of catch and inventory data the assessment of production became very difficult. By creating various thematic map based on biomass, salinity, mesh/Gear, craft and historical catch data its became easy to analyze and predict the production. Land use pattern, change in land use pattern, sediment deposition all can be monitored and quantified with the application RS and GIS. Even though this technology were used by Carrie Smith and others for linkage between change in estuarine vegetation and salinity. They found GIS and aerial photographs are very much useful for mapping of estuarine vegetation.

GIS is useful in assembling of wide ranges of data sources. Readily available sources of GIS data such as Coastline political boundaries are frequently used data. Many time Census data also used in GIS for analysis. Land use / land cover is either obtained from previous compilation or interpret from satellite data. Catchments boundaries based on DEM or delineated from relief maps. GIS also use available data to generate Information like rainfall and runoff.

Water body information system of West Bengal that was developed at CIFRI also used the RS data (LISS III) and GIS and found for preparing inventory of water bodies RS data and GIS are very much useful. This information system able to give the spatial information (latitude longitude nearest village block district water area and shape) of water bodies which have the water area more then 10 ha.

14.7 Monitoring of mangrove forest

Mangrove forests form one of the primary coastal ecosystems in the tropical and subtropical region of the world. They are biologically diverse and have therefore traditionally been utilized for food resources, firewood, charcoal, timber and other minor products. However, mangrove ecosystems are very sensitive and fragile. Deatruction of mangrove forest and change in species composition directly shows the impact on biodiversity. In India and abroad various scientific groups are using GIS remote sensing and aerial photographs for estimation and monitoring of mangrove forest. Yosif ali Hussain and others utilized the RS data (Landsat MSS, SPOT- XS and Landsat TM) and GIS for monitoring of mangrove forest of Indonesia (the delta of Mahakam River). Their study reveals that it is possible to monitor mangrove deforestation with reasonable accuracy using optical and radar satellite images and GIS. In India Dr Sridhar Vadlapudi applied GIS and RS data for Identification and quantification of changes in mangrove forest near Kakinada bay and Atanu Kumar Raha applied same techniques for monitoring change in Sunderban Mangrove forest. Both of them found that GIS and RS data are very much useful monitoring and quantification of change in mangrove forest area. The same type studies were also conducted in Thailand and Vietnam and they found GIS is very much useful in Coastal zone environment management.

14.8 Aquaculture site selection and development

Ross and others were used GIS for site selection for salmonid cage aquaculture in small bay. This study used a sequential screening process for different criteria to identified potential aquaculture site for cage culture. Depth, current velocity salinity dissolved oxygen, temperature and other factors (such as Infrastructure, topography and exposure) were used as simple overlay layer. Independent maps were prepared for above parameters using base topographic map for depicting their spatial characteristics. Different scores were assigned for parameters according to suitability for culture like depth less then 6m as 0 and 1 for more then 6 m similarly. They also shows that GIS can use to address biophysical, economic and social factors for decision support system.

14.9 Brackish Water Aquaculture Site Selection

Brackish water aquaculture has tremendous potential due to ever increasing demand of prawns and export earnings. Aquaculture development and planning require comprehensive data on land use and water resources. The land use/wetland information has been used for evaluating quality of the surrounding coastal waters, saltpans, aquaculture ponds for feasibility of brackish water aquaculture. The areas under mangroves/marsh are of vital/critical concern and were avoided for site selection.

The spatial distribution of mud/ tidal flat areas, which are most suitable from substrate condition point of view give an idea about potential available for brackish water aquaculture. IRS II data was used for preparing coastal land use maps by Space application center on 1: 50000 scale. This map shows various features like forest, lakes/ponds, mudflats, river, stream, High water line, Low water line, Mangrooves, etc. These maps along with other engineering, biological, meteorological and socio-economic related parameters were integrated using GIS for evaluating site suitability. Site evaluation is carried out to determine whether a site is suitable or not and to appreciate about requirements to make the site suitable. This procedure had ensured development for brackish water aquaculture with minimum damage to the ecology of the area.

14.10 Web GIS in Fisheries

Web-based GIS is becoming more and more common as time passes. GIS for anyone, anywhere, anytime - the Web enables geographic information to be shared and distributed quickly and cost effectively. Web GIS is a integration of Word Wide Web (WWW) and GIS technologies. In Web GIS, GIS server prepares maps and launches on web sites. In this technology, there are two things WebGIS server and map browser. Map browser sends the request to GIS server and GIS server creates the map and sends to the client. It is very useful tool for dissemination of information (Map based information) in large no of clients.

15. Summery

The application of remote sensing is not now a new tool for mapping of wetland and other natural resources. Previously the spatial resolution was very low, so the mapping of small water bodies was not possible. But after lunching of new satellite (OrbView and IKONOS, etc.) with high spatial resolution the mapping of water body less than 0.05 ha is possible. There are few parameters studied with the remote sensing but in future some more parameters can be correlated. Hyperspectral data now available from several satellites, allow better discrimination between pigments, thus allowing the identification of broad algal groups. Such information allows us to move away from empirical approaches now being used and develop algorithms that will allow us to use the remote sensing for assessment, monitoring and management of fisheries resources.

A GIS is computer-based tool for geographical analysis of information. It is not simply a digitized map, nor does it hold maps. It holds a database of spatial data and attribute of descriptive information about features on a map, which can be used to create desired maps. GIS has a significant role to play in the decision making process in fisheries. At various levels i.e. Field, regional, national and global levels. This is the one of the important tool of information technology, which is very useful to fishery scientist/ Managers to arrive at better decisions by integrating research results with intensive location specific information on various biophysical and socioeconomic variables. This technology allows to use and handle a wider range of spatial database such as soil, hydrology, crop, fisheries, weather etc. and integrate with socioeconomic variable and with the new technologies that are being developed from time to time. Potential areas for growth in GIS application to fisheries management include real time analysis, stock assessment, spatial demand and supply analysis, inventory management, environmental impact assessment biodiversity degradation assessment and disaster management.

Current information technology trends and future GIS tools will provide a range of function embedded in various components that can be tailored for specific use. GIS enable application by incorporating software components supporting geographic manipulation and visualization that are needed for specific application for development of capture fisheries aqua culture development wetlands management reservoirs management etc.

FUNDAMENTALS OF THE INTERNET AND THE WORLD WIDE WEB

C. Chandra Shekhar[1]

Abstract

The paper explores the evolution of the Internet to what it is today and how simultaneously websites have transformed from static electronic brochures to interactive information kiosks.

1. Evolution of the Internet

You may associate the Internet with electronic mail and business applications — but it came into being with global thermonuclear war in mind.

The Internet had its genesis more than a quarter of a century ago in the United States Department of Defense's Advanced Research Project Agency (ARPA). It was envisioned as an experiment to explore the possibility of a disaster-proof nationwide computer system, where military personnel could share messages and data no matter where they were or what had happened. It would alleviate the risk of a centralized control station which could be knocked out, thus shutting down communication altogether.

In the late 1960s, U.S. government-funded researchers in California working on the experiment explored ways to allow computers to share information. At that time, computers were almost totally unable to electronically communicate, and to move data from one computer to another required physically transporting punch cards or reels of tape.

In 1969, the research group connected computers at the University of California Los Angeles, the Stanford Research Institute in Menlo Park, California, the University of California Santa Barbara, and the University of Utah, using what was then a new networking technology. It enabled computers previously separated by technological and computer language barriers to share information with each other. It was the computer equivalent of being able to call across town and talk to a friend instead of having to leave home, travel there and bring a translator. The network was named ARPA Network, or "ARPANet."

The network soon began to expand, with more and more university sites coming on-line across the country, and the technology of how to establish links spreading. Along with the communication link, researchers had made progress on the component considered seminal to the Defense Department: making the network resistant or impervious to intentional disabling.

[1] **Correspondence:** C. Chardrasekhar, PGDM (XIMB) Managing Director, Catalyst IT Solutions Pvt. Lmt. www.citspl.com

The upshot was the Internet Protocol (IP), which enabled any number of computer networks to link up and act as one — and eventually gave the Internet its name. The IP meant that the communication network between the computers was not dependent on any single computer to operate — and could not be shut down through destroying one or even several of the computers. Years later, this principle was put to the test in the Gulf War, when Desert Storm forces sought to wipe out the Iraqi computer system — which was linked through the Internet — and were unable to do so.

Thus, the Defense Department got what it wanted. And a new, growing system of computer networks was spawned.

The Internet's growth was slow but steady. Author Jerry Pournelle, who first got an ARPANet account in 1977, recalled the infancy of the Internet. Back then, when a convention featured a party for network users, only a few dozen people would show. *"In 1978, after a year or so of experience with ARPANet, I predicted that by the year 2000 anyone in the civilized world would be able to get the answer to any question whose answer is known or calculable,"* said Pournelle.

That prediction is right on target. "On the other hand," Pournelle said, "in 1978, few foresaw the explosive growth of the microcomputer industry or the wild proliferation of network connections."

While the Defense Department continued to pay for just about every aspect of the network, in the early 1980s the military network split off from the educational networks and became known as MILNET. The National Science Foundation (NSF) became interested in the Internet, formed its own network on the Internet, NSFNet, and decided to disseminate the Internet Protocol technology to universities nationally. Concurrently, other government agencies, such as the National Aeronautics and Space Administration (NASA) and the Department of Energy, got involved. The Internet soon was crowded with government agencies, and a broad, powerful but chaotic system was born.

In time, NSF's network became much more dominant than ARPANet, and the vast majority of new as well as old Internet business was coming to NSFNet. By 1990, ARPANet shut down; but the NSFNet, despite having outlived its original research-oriented purpose, was so entrenched in the Internet that it lived on, maintaining and expanding the Internet.

Also, as years passed, the scientific presence and interest was totally drowned by waves of users seeking access for information, communication, entertainment and commercial enterprise under the NSFNet "acceptable use" policy which afforded wide public access to the network. In 1994, NSFNet wound down, and its traffic is being taken over by commercial networks — Merit Network, IBM and MCI.

Today, although far from its original concept, the Internet satisfies far more needs and wants than ever envisioned. It is a true example of technology that can help mankind move into the future.

2. The World Wide Web

It was 20 years ago this month that Tim Berners-Lee wrote up a proposal for his bosses at CERN, the European Laboratory for Particle Physics, for a hypertext-based system of connecting information that would use the TCP/IP protocols of the Internet plus a simple addressing scheme. Of course, today, billions of people use the system that came out of that proposal: We call it the World Wide Web, or more familiarly, the Web.

'Vague, but exciting', were the words that Mike Sendall, Tim Berners-Lee's boss wrote on the proposal for an information management system, allowing Berners-Lee to continue.

(You can read the full text of the proposal at "http://www.w3.org/History/1989/proposal.html")

Berners-Lee readily admits that his work is built on lots of previous innovations. The concept of hyptertext goes back a long way, to Vannevar Bush's Memex proposal and Ted Nelson's Xanadu project. The Internet is built on pioneering efforts by people such as Donald Davis and Paul Barran, who did early work on packet switching, and Vint Cerf and Robert Kahn, who created the TCP/IP protocol, along with many others. Berners-Lee's big contribution was marrying hypertext and the Internet together in a simple way.

Even once that now-obvious concept was developed, it took a long time for the Web to mature. Berners-Lee's proposal was actually shelved then, and it wasn't until the fall of 1990 that he started coding the program called World Wide Web and had developed the basic Hypertext Transfer Protocol (HTTP), Hyptertext Markup Language (HTML), and Universal Resource Identifier (URI, later URL) that are still the basic building blocks of the Web.

And while Berners-Lee developed a basic browser, it took Mark Andreesen and Eric Bina of the National Center for Supercomputing Applications at the University of Illinois to create Mosaic, the first modern Web browser to gain a large audience. Mosaic, in turn, led to Netscape, and the rest is history.

www

The letters "www" are commonly found at the beginning of Web addresses because of the long-standing practice of naming Internet hosts (servers) according to the services they provide. So for example, the host name for a Web server is often "www"; for an FTP server, "ftp"; and for a USENET news server, "news" or "nntp" (after the news protocol NNTP). These host names appear as DNS subdomain names, as in "www.example.com".

This use of such prefixes is not required by any technical standard; indeed, the first Web server was at "nxoc01.cern.ch",[42] and even today many Web sites exist without a "www" prefix. The "www" prefix has no meaning in the way the main Web site is shown. The "www" prefix is simply one choice for a Web site's host name.

However, some website addresses require the www. prefix, and if typed without one, won't work; there are also some which must be typed without the prefix. Sites that do not have Host Headers properly setup are the cause of this. Some hosting companies do not set up a www or @ A record in the web server configuration and/or at the DNS server level.

3. Evolution of Website Content and Design

Web 1.0 is a retronym which refers to the state of the World Wide Web, and any website design style used before the advent of the Web 2.0 phenomenon. It is the general term that has been created to describe the Web before the 'bursting of the dot-com bubble' in 2001, which is seen by many as a turning point for the internet.
It is easiest to formulate a sense of the term Web 1.0 when it is used in relation to the term Web 2.0, to compare the two and offer examples of each.

Terry Flew, in his 3rd Edition of New Media described what he believed to characterize the differences between Web 1.0 and Web 2.0.
"move from personal websites to blogs and blog site aggregation, from publishing to participation, from web content as the outcome of large up-front investment to an ongoing and interactive process, and from content management systems to links based on tagging (folksonomy)"

Flew believed it to be the above factors that form the basic change in trends that resulted in the onset of the Web 2.0 craze.

The shift from Web 1.0 to Web 2.0 can be seen as a result of technological refinements, which included such adaptations as "broadband, improved browsers, and Ajax, to the rise of Flash application platforms and the mass development of wigetization, such as Flickr and YouTube badges".

As well as such adjustments to the internet, the shift from Web 1.0 to Web 2.0 is a direct result of the change in the behaviour of those who use the World Wide Web. Web 1.0 trends included worries over privacy concerns resulting in a one-way flow of information, through websites which contained 'read-only'.material. Widespread computer illiteracy and slow internet connections added to the restrictions of the internet, which characterised Web 1.0. Now, during Web 2.0, the use of the Web can be characterized as the decentralization of website content, which is now generated from the 'bottom-up', with many users being contributors and producers of information, as well as the traditional consumers.

To take an example from above, Personal web pages were common in Web 1.0, and these consisted of mainly static pages hosted on free hosting services such as Geocities. Nowadays, dynamically generated blogs and social networking profiles, such as Myspace and Facebook, are more popular, allowing for readers to comment on posts in a way that was not available during Web 1.0.

At the Technet Summit in November 2006, Reed Hastings, founder and CEO of Netflix, stated a simple formula for defining the phases of the Web:
" Web 1.0 was dial-up, 50K average bandwidth, Web 2.0 is an average 1 megabit of bandwidth and Web 3.0 will be 10 megabits of bandwidth all the time, which

will be the full video Web, and that will feel like Web 3.0".
Some typical design elements of a Web 1.0 site include:

- Static pages instead of dynamic user-generated content.
- The use of framesets.
- Proprietary HTML extensions such as the <blink> and <marquee> tags introduced during the first browser war.
- Online guestbooks.
- GIF buttons, typically 88x31 pixels in size promoting web browsers and other products.
- HTML forms sent via email. A user would fill in a form, and upon clicking submit their email client would attempt to send an email containing the form's details.

The term "Web 2.0" refers to a perceived second generation of web development and design, that aims to facilitate communication, secure information sharing, interoperability, and collaboration on the World Wide Web. Web 2.0 concepts have led to the development and evolution of web-based communities, hosted services, and applications; such as social-networking sites, video-sharing sites, wikis, blogs, and folksonomies.

The term was first used by Dale Dougherty and Craig Cline and shortly after became notable after the O'Reilly Media Web 2.0 conference in 2004. Although the term suggests a new version of the World Wide Web, it does not refer to an update to any technical specifications, but rather to changes in the ways software developers and end-users utilize the Web. According to Tim O'Reilly:
" *Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as a platform, and an attempt to understand the rules for success on that new platform. "*

O'Reilly has said that the "2.0" refers to the historical context of web businesses "coming back" after the 2001 collapse of the dot-com bubble, in addition to the distinguishing characteristics of the projects that survived the bust or thrived thereafter.

However, the argument exists that "Web 2.0" does not represent a new version of the World Wide Web at all, but merely continues to use so-called "Web 1.0" technologies and concepts. Techniques such as AJAX do not replace underlying protocols like HTTP, but add an additional layer of abstraction on top of them. Many of the ideas of Web 2.0 had already been featured in implementations on networked systems well before the term "Web 2.0" emerged. Amazon.com, for instance, has allowed users to write reviews and consumer guides since its launch in 1995, in a form of self-publishing. Amazon also opened its API to outside developers in 2002. Previous developments also came from research in computer-supported collaborative learning and computer-supported cooperative work and from established products like Lotus Notes and Lotus Domino.

In a podcast interview, Tim Berners-Lee described the term "Web 2.0" as a "piece of jargon":

"Nobody really knows what it means...If Web 2.0 for you is blogs and wikis, then that is people to people. But that was what the Web was supposed to be all along."

Other criticism has included the term "a second bubble" (referring to the Dot-com bubble of circa 1995–2001), suggesting that too many Web 2.0 companies attempt to develop the same product with a lack of business models. The Economist has written of "Bubble 2.0". Venture capitalist Josh Kopelman noted that Web 2.0 had excited only 530,651 people (the number of subscribers at that time to TechCrunch, a Weblog covering Web 2.0 matters), too few users to make them an economically viable target for consumer applications. Although Bruce Sterling reports he's a fan of Web 2.0, he thinks it is now dead as a rallying concept.

Critics have cited the language used to describe the hype cycle of Web 2.0 as an example of Techno-utopianist rhetoric. Web 2.0 is not the first example of communication creating a false, hyper-inflated sense of the value of technology and its impact on culture. The dot com boom and subsequent bust in 2000 was a culmination of rhetoric of the technological sublime in terms that would later make their way into Web 2.0 jargon. Indeed, several years before the dot com stock market crash the then-Federal Reserve chairman Alan Greenspan described the run up of stock values as irrational exuberance. Shortly before the crash of 2000 a book by Shiller, Robert J. Irrational Exuberance. Princeton, NJ: Princeton University Press, 2000. was released detailing the overly optimistic euphoria of the dot com industry. The book Wikinomics: How Mass Collaboration Changes Everything (2006) even goes as far as to quote critics of the value of Web 2.0 in an attempt to acknowledge that hyperinflated expectations exist, but that Web 2.0 is really different.

Berners-Lee however, has been focused mostly on building what he calls "the Semantic Web"--on which a computer program can actually process a Web page enough to get some meaning from it. He's been talking about this idea for years, and a number of other people are working on it as well, including the World Wide Web Consortium, which he runs; the Web Science Research Institute, and people in education and in private companies. Some of these Semantic Web ideas are also referred to as "Web 3.0."

But Web 3.0, Web 2.0, and even the classic Web sites we're all used to wouldn't exist--at least not in their current forms--if it wasn't for that original vision – the World Wide Web.

STATISTICAL PROGRAMS IN QUANTITATIVE GENETICS AND SELECTIVE BREEDING STUDIES

Kanta Das Mahapatra[1]

1. Introduction

India possesses rich fish germplasm resources, accounting for about one-tenth of the 20,000 and odd species of fish known in the world. The Ganga network of rivers in the North, the Brahmaputra in the East, the Sutlej, the Narmada & the Tapti in the West and the Mahanadi, the Godavari, the Krishna and the Cauvery in the South are very rich sources, harbouring bulk of the important fish fauna.

Apart from the already existing different aquaculture practices in the country such as extensive, intensive and semi-intensive farming systems, the availability of huge quantities of organic wastes from plants and animals gives scope for integrated fish farming.

Exploitation of genetic potentials of carps has been initiated seriously from early 1980s particularly through genome manipulations and later through selective breeding from 1990s, to add to the already developed culture technologies, for further enhancement of fish production in the country.

Selective breeding in increasing production level is well established in agriculture and animal husbandry. Today the high yielding crops and land animals are totally depending on genetically improved domesticated breeds. This has not been true for aquaculture. Proper exploitation and utilization of genetic potential is lacking in aquaculture. Less than 5% of the total output of the aquaculture production is coming from improved breeding progamme. Aquaculture species are thus genetically much closer to their wild counter part than the land animals and plant species. During the last few years it has been well documented that high selection response can be obtained in fish as well as in shell fish for economic important traits like growth, disease resistance, flesh quality etc.

Carp occupies very important position in freshwater aquaculture. 95% of world production of carps is coming from Asia only. Annual growth rate of carps is 11% and most of the low-income groups consume the carps. Very little genetic improvement work has been carried out in carps. Targets for genetic improvement of carps is complex because of the diversity of species, farming system, socioeconomic scenarios and longer generation period. However, various commercially important traits can be improved through selective breeding in carps.

Different statistical program can be used to analyze data in an effective manner. Some of the programs are discussed here which can be applied in quantitative

[1] **Correspondence**: Kanta Das Mahapatrra, Principal Scientist, *Central Institute of Fresh Water Aquaculture, Kausalyaganga, Buubaneswar-751002, Orissa: Email* kdmahapatra@yahoo.co.in

genetics study to determine different parameters used in the selective breeding programme.

2. Selective Breeding

The potentials of genetics have already exhibited promising trends in aquaculture too. The correct breeding procedures followed through selective breeding in the case of Atlantic salmon and rainbow trout in Norway, Channel catfish in USA and Nile tilapia in Philippines are the standing examples in this regard (Gjedrem, 1997). The recent selective breeding programme of rohu (*Labeo rohita*) in India too is another example of the kind (Reddy *et al.*, 2002).

Before success of induced breeding through hypophysation in late fifties rivers are the main source for seed collection of Indian major Carps. With the introduction of induced breeding technique, hatcheries have been able to produce enough quantity of carp seed. The hatcheries in India hardly follow any genetic norms to produce carp seed. A limited number of brood fishes are used repeatedly for successive generations. As a result the quality of carp seed is showing negative effect of inbreeding i.e. slow growth rate, disease proneness etc.

Inbreeding

Inbreeding occurs due to mating of closely related individuals. Genetically inbreeding leads to homozygosity. Almost all individuals carry deleterious recessive genes, which are hidden in heterozygous state. Related individuals are likely to share common genes and probability of pairing of deleterious recessive genes gets enhanced with the increase of closeness between the parents. Due to which inbreeding depression occurs to the population with decrease growth efficiency, disease resistance and survival.

Effective population size

Effective population size is one of the most important concepts in the management of a population in that it gives an indication about the genetic stability of the population. It depends upon several factors such as total number of breeding individuals, sex ratio, mating system and variance of family size. The effective population size can be calculated by following formula,

$$N_e = 4N_f X N_m / N_f + N_m$$

Where N_e = Effective population size

N_f = Number of female brood fishes used for seed production

N_m = Number of male brood fishes used for seed production

Effective population size is inversely related to inbreeding

$$\Delta F = 1/2N_e$$

So $\Delta F = 1/8N_f + 1/8N_m$

Where ΔF= Rate of inbreeding per generation
N_e= Effective population size
N_f = Number of female brood fishes used for seed production
N_m= Number of male brood fishes used for seed production

So to improve the genetic status of any population, hatchery managers should decrease inbreeding rate and increase effective population size. Selective breeding plays vital role to improve genetic status of fish in a positive direction which has already been demonstrated in case of Salmon at Norway, Tilapia at Philippines and Rohu at CIFA, India.

Objectives of selective breeding

Objectives of a selective breeding programme is to change the average performance of the targeted trait i.e. growth rate, disease resistance, better flesh quality, feed conversion efficiency etc. of the population in a favorable direction.

3. Estimation of phenotypic and genetic parameters

Estimation of phenotypic and genetic parameters is essential in the design of genetic improvement programs. They include phenotypic and genetic variances and covariances and their linear functions such as heritability and phenotypic and genetic correlations. They enable the prediction of genetic gain and of correlated responses in genetic improvement program.

Selection is an age-old process in nature. Fittest organism survives and other eliminated. Selection also can be achieved artificially. In this process best individuals are selected as parents so that parents pass on their superior genes to their progeny and better progeny can be obtained. Selective breeding based on principle of quantitative genetics. Which indicated that that phenotype of an individual, which can be measured or scored, could be partitioned in to two components. One attributable to the influence of genotype i.e. the particular assemblage of genes possessed by the individual and other one attributes to the influence of environment i.e. all non-genetic components.

So $P = G + E$
Where P = Phenotype of an individual
G = Genotype
E = Environmental (Non-genetic) component

Quantitative phenotype exhibit continuous variation, the only way to study them is to analyze the variance that exists in a population. The phenotypic variance (V_P) that is observed for a quantitative trait is the sum of the genetic variance (V_G) and (V_E) and the interaction that exists between the genetic and environmental variance (V_{G+E}).

So $V_P = V_G + V_E + V_{G+E}$

Genetic variance is the component of interest in selective breeding program. V_G is further subdivided in to three component i.e. additive genetic variance (V_A), dominance genetic variance (V_D) and the epistatic genetic variance (V_I).

$V_G = V_A + V_D + V_I$

V_A, V_D and V_I differ from each other according to their mode of inheritance. Dominance genetic variance is the variance that is due to the interaction of the alleles at each locus. Because of this, V_D can not be inherited; it is created a new in each generation. Since it is the interaction between alleles at each locus so V_D is a function of the diploid state as alleles occur in pairs. During meiosis homologous chromosomes and allelic pairs are separated during reduction division and the chromosome complement reduced to half. So V_D is not transmitted to next generation.

Epistatic genetic variance (V_I) is due to interaction of alleles between two or more loci. Epistatic interaction occurs across loci so it is also not transmitted from parents to offspring. This also created a new in each generation.

V_A or additive genetic variance is the additive effect of genes. It is the sum of the effects of each allele that is responsible for phenotype. It does not depend on specific interaction or combination of alleles so it is not disrupted due to meiosis so additive genetic variance transmitted from parents to offspring. It is transmitted in a reliable and predictable manner. V_A is also called the variance of breeding value.

Heritability

Heritability describes genetic component that is not disrupted by meiosis. The proportion amount of phenotypic variance (V_P) that is controlled by V_A is called heritability (h^2).

$$h^2 = V_A / V_P$$

Once you know the heritability response of selection can be predicted as

$$R = S * h^2$$

Where R= response to selection,

S = selection differential (difference between offspring and parent generation)

h^2 = heritability of the trait

Selection methods

Several selection methods are available for obtaining additive genetic variance. The methods differ with respect to which type of relatives that provide information used for the selection decisions. The objective of all the methods is to maximize the probability of correct ranking of animals with respect to their breeding value, an estimate of each individual ability for producing high / low performing offspring. The breeding value of an individual cannot be estimated on basis of the phenotypic value of the trait (s). In fish such records are usually obtained from the individual itself (Individual selection), full and half sibs (family selection) or from all three sources of information (combined selection).

Breeding objectives and selection indices

Selection indices have two main uses such as it combines information from different relatives for population trait and it combines the genetic merit for different traits in an aggregate breeding value. The information regarding different traits may vary widely. Some coming from an animal relative and some from the animals own performance for traits which, are expressed once or repeatedly during its life. These factors make wise selection a complicated and uncertain procedure. In addition, fluctuating, vague and sometimes erroneous ideals often cause the improvement resulting from selection to be much less than could achieved if these obstacles were overcome.

4. Steps in designing breeding programme

Following steps may be considered in designing breeding program

- Production system
- Breeding system and breeds
- Formulation of objectives of the system and calculation of economic values (weight)
- Development of selection criteria and estimation of selection parameters
- Design of animal evaluation system
- Design mating for selected animals
- Design system for expansion
- Monitoring and comparing alternative programs.

The most important decision in any animal improvement program is the choice of breeding objective. A badly chosen objective will lead the genetic progress in the wrong direction. There may be conflicts to the fragmented nature of production chain. Interest of the breeders, multiplier and producers may differ. The breeding objectives or the aggregate genotype is about where to go. But the selection index is about how to get there. The selection index is a linear function of recorded characters used as a basis for selection. It is derived so that the correlation between the index and the aggregate genotype are maximized. Such a linear function includes characters of the animals and its relatives to predict and improve the aggregate genotype. The overall economic merit can be thought of as a breeding value of a new trait.

While developing breeding objectives, the starting point should be the determination of an equation for economic benefit derived from a population. Genetic improvement typically takes place in a very small fraction of the population, which is multiplied and disseminated, to the production systems. Breeding objectives should be defined for the production system. Instead of using specific relatives, parameters may be estimated from the complete pedigree.

Trait of selection can be of four types i.e.
Production traits
Morphometric traits
Reproductive traits
Quality traits

Among all traits growth rate occupies the top in the list of preference.

5. Role of population size in selection programs

Some of the considerations we often make about genetic improvement assume we are working with populations of infinite size but in reality we are always working with population of finite size. If the population size is of finite size then rate selection response declines, the maximum response is reduced and the selection limit is reached faster. The inbreeding coefficient increases and that may have undesirable effects. Rate of inbreeding can be calculated from effective population size.

$F=1/2$ Ne after 't' generations $F_t = 1-(1-(1/2Ne)t$

Response per generation $R= I\ h\ \sigma_A$

Under inbreeding σ_A is reduced by

$\sigma_{A---F} = \sigma_A\ (1-F)^{0.5}$

Aim of any selection is to increase frequency of favorable alleles. Chance of missing out in favorable alleles increases with reduced Ne. When all are fixed, we are at limit. It is accepted that rapid inbreeding is more harmful than gradual mild inbreeding. Incidence of genetic defect increases. For a rate of inbreeding of ~ 1% per generation 50 brood fish pairs or more are required.

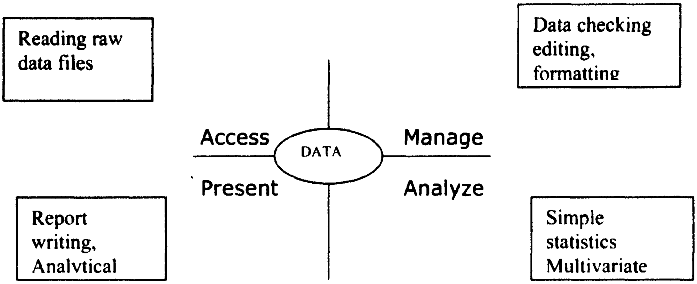
For analyzing all the data for the selective breeding program different statistical programs are used. The accuracy of any breeding program depends on the accuracy of analysis of genetic merit of the individuals. Some of the statistical programs are described below.

6. Statistical Analytical System (SAS)

SAS software is a combination of a statistical package, a data base management system and a high level programming language. It can run on a wide variety of computers and operating systems. SAS programs communicate with the computer by SAS ststements. It can be used to describe a collection data and produce a variety of statistical analyses. SAS package has following functions and advantages.

- It is integrated suite of modular products designed to meet a range of focused information needs.
- This software targets and solves industry specific problems.
- Allows to easily performing virtually any type of statistical analysis.

- Comprehensive to perform the most sophisticated multivariate analysis
- Functional and built around the 4 data-driven tasks crucial to every application



SAS program has 2 kinds of steps

1) Data step (used to create SAS data set)
2) Proc step (Used to process SAS data set)

SAS data set can also be created from excel or dbase file through import and export menu. Immediately following the data is a series of Procs. They perform various functions and computation on SAS data sets. Since we want a list of subjects and scores in subject order, we first include a sort procedure by line. Sorting can be multilevel if desired.

SAS program understands the keywords TITLE, ID and VAR and interprets what follows in proper context. Each statement ends with semicolon (;). A semicolon in a SAS program is like a period in English. Probably most common error found in SAS programs is omission of the semicolon. Through proc means procedure mean Number of observation, Standard deviation, minimum and maximum value comes as default. But along with proc means statement CV (Coefficient of variation), maxdec= 1/2/3, sterr (standard error) statement can also be given.

SAS programs are like sandwiches. They began with DATA statements (Bread), which tells about the data set. Then comes the data proper (middle part of sandwiches). Finally the PROC statement (bread) which specify the analyses to be performed.

When the program executed, it produces something called the SAS LOG and the SAS OUTPUT.Any SAS error messages will be found in the SAS LOG. Solution of the data analysis can be observed in the SAS output.

Use of SAS in Rohu selective breeding program-case study

In the rohu selective breeding program, individual tagging was done to identify fishes of different fullsib families in the communal pond. After fishes attained taggable size i.e. 10-15 g they were tagged with Passive Integrated Transponder (PIT) tags and stocked in communal rearing ponds for further grow out experiment. Before tagging, initial body weight of fishes was noted and during final sampling also final individual body weights were noted.

7. Adjustment of fixed effects

All recorded data need to be carefully edited to eliminate errors during recording. Special attention was paid to the outliers (too small or too large value for trait or an ID) and individual having same ID.

As selection index procedure is used to calculate breeding value, the data need to be preadjusted for fixed effects i.e. pond. Multiplicative adjustment is applied in the rohu-breeding program.

Estimation of breeding value

After adjustment of pond effect, corrected body weights are being considered for the estimation of breeding values. It is estimated using following formula

$$\hat{I} = b_1(P_i-P_{fs})+b_2(P_{fs}-P_{hs})+b_3(P_{hs}-P_{pop})$$

Where

$\hat{I}$ = Predicted breeding value

P_i = Adjusted phenotypic record on the individual

P_{fs} = Adjusted phenotypic mean of n-fullsibs, P_i included in P_{fs}

P_{hs} = Adjusted phenotypic mean of mn halfsibs, P_i included in P_{hs} and m is the number of dams nested to each sire.

P_{pop}= Adjusted population mean

b_1 b_2 b_3=the weight s given to each source of information

The mixed procedure

The mixed procedure fits mixed linear models (models with both fixed and random effects). A mixed model is a generalization of the standard linear model used in the GLM procedure. One can analyze the data with several sources of variation instead of just one.

Some features of MIXED model are as follows,

- Covariance structures, including simple random random effects, compound symmetry, unstructured data
- GLM type grammar using MODEL, RANDOM and REPEATED statements for model specification and CONTRAST, ESTIMATE AND LSMEANS statement for inferences
- Appropriate standard errors for all specified estimable linear combinations of fixed and random effects and corresponding t- and *F* test.

- Capacity to manage unbalanced data
- Ability to create a SAS data set corresponding to any printed table.

8. The inbreeding procedure

The inbreed procedure calculates the covariance or inbreeding coefficients for the pedigree. PROC INBREED is unique in that it handles vary large populations

The inbreed procedures has two modes of operation. One mode carries out analysis on the assumption that all the individuals belong to the same generation. The other mode divides the population into non-overlapping generations and analyzes each generation separately assuming that the parents of individuals in the current generation are defined in the previous generation. Proc inbreed also compute average of the covariance or inbreeding coefficient within sex categories if the gender of individuals is known.

9. Strain comparison and heterosis estimation

The diallel cross is an efficient design for evaluating the performance of purebred and crossbred stocks or lines. In complete diallel cross each stock is crossed with itself and all other stocks. For each species that may be stripped artificially the following mating procedure should be followed to obtain an equal contribution of genes from each cross and from each individual within a cross For each stock, milt from a given number of sires is used to fertilize with an equal number of eggs from a given number of dams of the same and of each of the other stocks. Thus both crosses between the pure stocks and all possible cross combinations between stocks including reciprocal are produced

A complete design may provide estimate of additional genetic effect of each stock, overall heterosis, general heterosis for each stock and specific heterosis for each cross. Following formula can also be used to determine above-mentioned aspects.

P_n = mean phenotype of all purebreds

$X_{n(n\ 1)}$ = mean phenotype of all crossbreds

Pj/i = mean phenotype of one purebred
= $P_n + gj/I + g^m\ j/i$ (here g^m is the maternal effect)

$h^1..$ = average heterosis for all crosses
$= \Sigma\ h_{ij} = X_{n(n\ 1)} - P_n$

h^1_{ij} = average heterosis for reciprocal cross
$= X_{ij} + X_{ji} - P_i - P_j)/2$

h^1_j = average strain heterosis
$= \Sigma h^1_{ij}/(n-1)$

$g^m_j = \Sigma\ (X_{ij} - X_{ji})\ /\ n$

g_{ij} = average individual effect of each strain
= $P_j + P_n - g^m_j$

10. Selection Index Program (SIP)

Selection index theory plays an important role in teaching animal breeding. It can be used to determine the expected response to selection. The department of Animal breeding and the department of informatics have worked on the development of user-friendly program to calculate selection indexes. Johan Van Arendonk developed the program. The program SIP can calculate many selection indexes. These selection indexes may be based on economic value as desired gain or on a combination of both. The program is written in Turbo vision. The program can create a number of temporary files, which can be deleted after leaving the program. The program will start in DOS promt. The important steps in operating SIP is as follows,
irst we have to set trait dialog box.

Then we have to set r_p and r_g boxes.

Next we have to set group dialog box which includes additive genetic relationship and number of observation.

Then we have to set genetic relation dialog box

Common environment box is the next to be filled in.

Sources of index is marked next

Breeding goal is set next

And finally through calculation box index can be calculated.

11. Genetic Parameter Estimation in Excel (GPEX)

GPEX runs in Excel 97 or higher versions. Size is about 60 MB for analysis of about 5000 observations. User is not required to have advanced programming skills. To some extent it could be a good alternative to using top of the line statistical packages in estimation of genetic parameters. It can handle unbalanced data and handle 48 full sib groups with maximum 150 progeny per full sib groups. It can calculate genetic parameter like intra-class correlation, additive genetic, environmental and phenotypic variance, covariance, correlation, and heritability. It can also predict breeding values based on calculated genetic parameters from data and accuracy also can be calculated.

PEST

Pest is a software package for multivariate prediction and estimation. It covers fixed, random and mixed model. Pest read raw data and translates class codes like month name integer or character identification of animal into internal representation. Identification can be up to 16 characters long. A parameter file contains commands to pest. It consists of input data description, the statistical model, output, data transformations etc. Pest is written in FORTRAN 77. It consists of around 15000 lines of source code in 250 routines. PEST provides very general and flexible strategies to solve mixed model equations. Its objective has not been to maximize memory requirements. Together with the large amount of program code PEST does require a certain minimum amount of processing capability.

ASREML

ASREML estimates variance component under a general mixed model by restricted maximum likelihood (REML). Its scope covers genetic, multivariate, repeated measures, spatial and multi-environment analyses. It uses the average information algorithm and sparse matrix technique to efficiently solve large mixed models. The user interface is basic and assumes a good understanding of the models that can be fitted; the results may need to be imported into another statistical / reporting program for further processing. ASREML enables limited testing of some fixed effect in the model. ASREML is available in complied form of MSDOS, Windows 95/ NT. We need an ASCII editor to prepare the data and parameter file before running ASREML. Base name is in the name of .as command file. Output file names are generated from the input name by changing file extension from .as to .ars which is primary output file summaries the data, iteration sequence, the final variance parameters and solutions for fixed effects. The pin is an input file required for predicting means and functions of the variance components when the P option is specific. . pvs is the report produced with P option. In ASREML blank space in the data set can be taken care by "*" mark.

12. Conclusion

Determination of breeding value i.e. genetic merit of the individual is very important in selective breeding studies. Success in selective breeding program depends on the correct ranking of individuals according to its genetic merits. Different statistical packages can be utilized for this. However, SAS proved to be most effective program for that. However, other programs like ASREML is also equally effective and can also be utilized for determining different parameters for selective breeding studies.

SUBTRACTIVE GENOMIC APPROACH TO IDENTIFY POTENTIAL DRUG TARGETS AND ACTIVE SITES IN *CHLAMYDOPHILA PNEUMONIAE*

S C Patnaik, E H K Reddy, Gyana R Satpathy[1]

Abstract

Whole genome sequences of the human pathogen *Chlamydophila pneumoniae* and four other strains of same species were analyzed to identify common drug targets. A substractive genomic approach is applied to identify Holliday junction DNA helicase RuvB as the common non-human homologous gene among these four strains. A three-dimensional model of the Holliday junction DNA helicase RuvB protein was generated with homology modelling. The protein is analysed for identification of suitable target sites.

Keywords: *Chlamydophila pneumoniae*, BLAST, MSA, homologous, homology modeling, active site

1. Introduction

Chlamydophila pneumoniae is a widespread obligate intracellular gram negative bacterium that causes upper and lower respiratory infections worldwide [Grayston et al., 1990]. In addition to acute infections, several chronic inflammatory diseases have been associated with *C. pneumoniae* infection. Increasing evidence implicates that a persistent lung infection caused by *C. pneumoniae* may contribute to the initiation, exacerbation and promotion of asthma symptoms [Hahn, 1999; Cazzola et al., 2004].

Identification of the target molecule inside various metabolic pathways is the first step when designing drugs against a pathogen. The vastness of the pathogen genomes has made this work more difficult involving considarable sum of labor and time. As microbial populations have rapid growth rates, evolution of resistance can occur in relatively short time frames. Moreover, resistance genes can be dispersed rapidly by genetic exchange systems among diverse bacterial species. It is therefore essential to develop rapid processes to identify novel antibiotics. Genes that are conserved in different organism often turn out to be essential. These essential genes should not have any well-conserved homolog in the human host. Inactivation of essential genes by any drug should result in the lethal phenotype in the pathogen [Judson and Mekalanos, 2000].

Large amount of sequence data is generated from various microbial genome sequencing projects around the world. These datasets create a major challenge in the post-genomic era. The strategies for drug design and development are increasingly shifting from the genetic approach to the genomic approach [Galperin and Koonin, 1999]. Subtractive genomics has been successfully used by authors to locate novel drug targets in *Pseudomonas aeruginosa*, *Helicobactor*

[1]**Correspondence**: Gyana Ranjan Satpathy, Coordinator, Bioinformatics Centre, National Institute of Technology, Rourkela-769008, India: Email: gyanasatpathy@gmail.com

pylori [Sakharkar et al., 2004; Dutta et al., 2006]. The work has been effectively complemented with the compilation of the Database of Essential Genes (DEG) for a number of pathogenic micro-organims[Zhang et al., 2004].

2. Materials and methods

2.1 Identification Novel Drug Targets: Whole genome sequences of four strains of *C. pneumoniae* (*C. pneumoniae* AR39, *C. pneumoniae* J138, *C. pneumoniae* TW1839 *and C. pneumoniae* CWl029) were downloaded from the National Center for Biotechnology Information (NCBI) (ftp://ftp.ncbi.nlm.nih.gov/genomes/bacteria/). The strains are having a circular genome with 1052-1112 predicted protein coding sequences [Kalman et al., 1999]. From the complete genome sequences data, the genes whose sequence length is greater than 100 amino acids were selected out. These selected genes were then subjected to BLASTX against the DEG database (http://tubic.tju.edu.cu.deg) to screen out essential genes. A random expectation value (*E*-value) cut-off of 10^{-100} and a minimum bit-score cut-off of 100 were used [Dutta *et al.,* 2006]. The screened essential genes of *C. pneumoniae* were then subjected to BLASTX against the human genome to identify the non-human homologous proteins in the bacteria. The homologous were excluded and the list of non-homologous was compiled. The identified essential non-human homologous proteins were then classified into different groups based on biological function with the help of the Swiss-Prot Protein Database (http://us.expasy.org/sprot). The classified essential and non-human homologous proteins within the same function group were further analyzed to find highly conserved proteins common to all four *C. pneumoniae* strains by using MSA (Multiple Sequence Alignment) in ClustalX [Thompson et al., 1997]. These proteins are considered as common drug target for all four *C. pneumoniae* strains. The flow chart of the process is shown in the Figure-1.

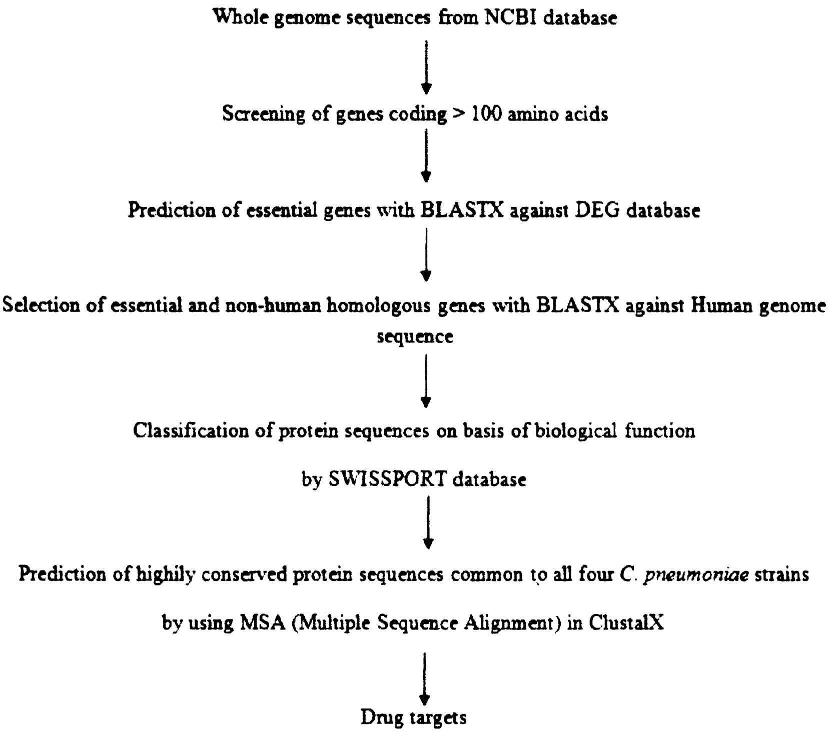


Figure 1: Insilico genomic approach for prediction of drug targets for *Chlamydophila pneumoniae.*

2.1 Homology Modeling: The essential and non-human homologous common protein sequence *i.e* Holliday junction DNA helicase RuvB protein was selected from *C. pneumoniae* strains as a drug target [Tuteja et al., 2006; Frick, 2003]. Homology modeling is usually the method of choice when a clear relationship of homology between the sequence of target protein and at least one known structure is found. This approach would give reasonable results based on the assumption that the tertiary structures of two proteins will be similar if their sequences are related and three-dimensional structure of proteins is better conserved during evolution than its sequence [Kroemer et al., 1996; Rost, 1999]. Building a homology model comprises four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These steps can be repeated until a satisfying modeling result is achieved. The query sequence Holliday junction DNA helicase RuvB was searched to find out the related protein sequences as a template by the BLAST program against the protein data bank (PDB). The sequence that showed maximum identity with high score, better resolution, better crystallographic R-factor and less E-value with the related family was aligned and used as a reference structure to build a 3D model for target protein.

The model was constructed by using the program Modeller9v2 under Windows. Modeller is a comparative protein structure modeling software. It is based on spatial restraints derived from the alignment and Probability Density Functions (PDFs) [Sali et al. 1993]. The 3D model of a protein is obtained by optimization of the molecular PDFs such that the model violates the input restraints as little as possible.

After building the protein 3D structure, in order to assess the overall stereo chemical quality of the modeled protein, Ramachandran plot analysis was performed using the program PROCHECK [Laskowski et al., 1993 ; Morris et al., 1992; Ramachandran et al., 1963]. Further evaluation of modeled structure was done by VERIFY3D [Eisenberg *et al.,* 1997], Errat [Colovos and Yeates, 1993], Prove [Pontius et al., 1996] and WHAT_IF [Vriend, 1990] through SAVS: structure analysis and verification server (http://nihserver.mbi.ucla.edu/SAVS/).

2.2 Active Site Identification: Sites of activity in proteins usually lie in cavities. The size and shape of protein cavities dictates the three-dimensional geometry of ligands that must fit like a hand in glove. The binding of a substrate typically serves as a mechanism for chemical modification or conformational change of protein [Binkowski et al., 2003]. Binding sites are often targeted by various ligands in attempts to interrupt related molecular processes. Active sites of a protein are a key factor for the flexible docking. Active sites of the target protein (Holliday junction DNA helicase RuvB) were predicted by using tool CASTp (computed atlas of surface topography of proteins) [Binkowski et al., 2003]. CASTp provides resources for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins. These include pockets located on protein surfaces and voids buried in the interior of proteins that are frequently associated with binding events. In addition, it measures the size of mouth openings of individual pockets for better accessibility of binding sites to various ligands and substrates.

3.1 Results and discussion

Whole genome sequences of the human pathogen *C. pneumoniae* were analyzed to identify drug targets. Strain wise distribution of genes is summarized in Figure 2. A total number of 4388 protein coding genes from four strains were studied. Out of which 3948 genes having more than 100 amino acids in their coding sequence were selected. This was on the assumption that proteins more than 100 amino acids long are known to be able to affect the catalytic activity of proteins and participate in protein complex formation which affect their enzyme activity [Yang et al., 1999].

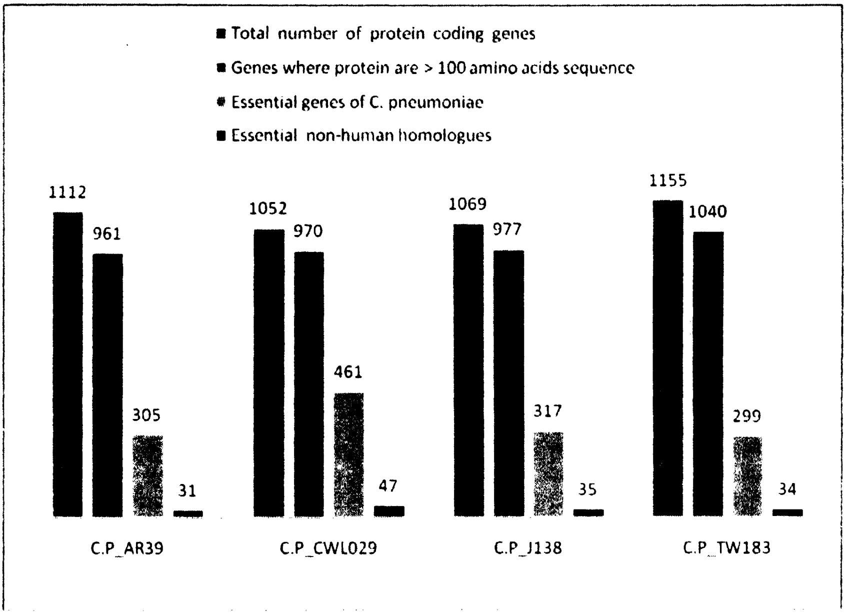


Figure 2: Gene distribution of four strains of Chlamydophila pneumoniae

After screening against the DEG database 1382 genes were identified as essential genes for four strains of *C. pneumoniae*. These genes are essential for survival of *C. pneumoniae*. Comparison of the screened essential genes with human genome resulted in identification of 147 genes which are essential and non-human homologous genes among four strains of *C. pneumoniae*. It was observed that though strains C.P_AR39, C.P_J138, C.P_TW183 show almost similar number of essential and non human homologous genes, the strain C.P_CWL029 shows a marked increase in number of respective genes.

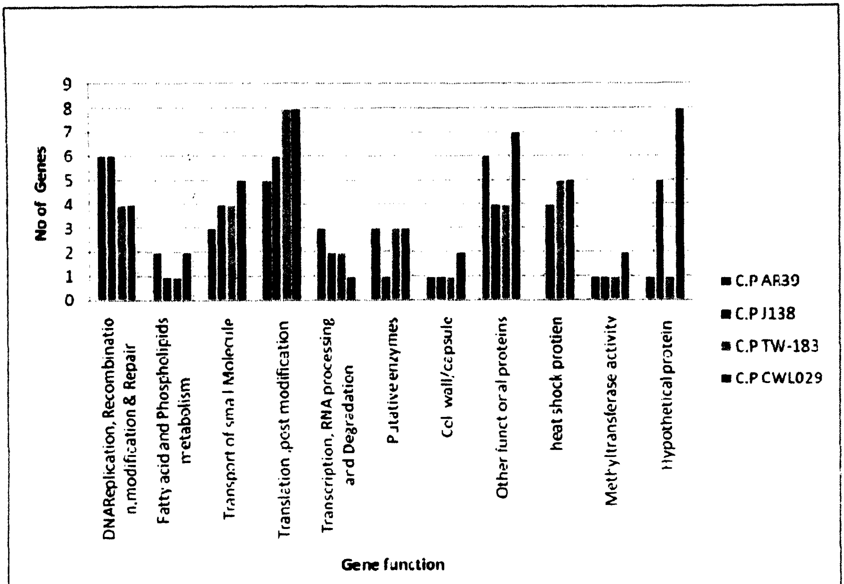


Figure 3: The graph showing non-human homologous essential genes encoding different proteins involved in a same biological function in comparison with four different strains.

These essential non-human homologous genes and their encoding protein were further categorized on the basis of the pathways involved in the basic survival mechanisms (Figure 3) such as: DNA replication, recombination, modification and repair, translation and post translation modification, transport of small molecule, transcription, RNA processing and degradation. These non-human homologous gene their encoding protein were represented as ideal drug targets, i.e. any disruption those genes may lead to bacterial death [Judson and Mekalanos, 2000]. These essential and non-human homologous genes cover 3-4% of total genome of the organism. The MSA shows seven different types of proteins are that are conserved among the strains and are essential non-human homologues. These seven genes can be considered as a novel drug target to design a drug Table 1. Among these seven genes Holliday junction DNA helicase RuvB was taken as a target for ligand binding study.

Table 1: The predicted drug targets of Chlamydophila pneumoniae

S. No	Protein name	Function of protein
1	Excinuclease ABC subunit A	DNA Replication, Recombination, modification and Repair
2	Holliday junction DNA helicase RuvB	DNA Replication, Recombination, modification and Repair
3	30S ribosomal protein S10	Translation ,post modification

4	30S ribosomal protein S2	Translation ,post modification
5	GTP-binding protein EngA	GTP-dependent binding, GTPase of unknown physiological role.
6	Acetyl glucosaminyl transferase	Acetylglucosaminyl transferase activity
7	Riboflavin-specific deaminase	Putative enzymes

Holliday junction is the central intermediate in homologous recombination which is formed as a result of a reciprocal exchange of DNA strands between two nearly identical DNA molecules. Homologous genetic recombination is essential in maintaining genomic stability such as to protect genomes from double-strand breakage or inter strand cross linkage. RuvA, RuvB and RuvC proteins form two complexes: a RuvAB complex with helicase which conducts branch migration activities and a RuvABC complex that resolves Holliday junction [West, 1997]. The RuvB hexamer is the chemo mechanical motor of the RuvAB complex. RuvB converts chemical energy from ATP into the dynamic force behind branch migration of Holliday junctions formed during DNA recombination and DNA replication. RuvA and RuvB bind to the four strand DNA structure formed in the Holliday junction intermediate, and migrate the strands through each other, using an assumed coiling mechanism. The binding of the RuvC protein to the RuvAB complex is thought to cleave the DNA strands, thereby resolving the Holliday junction. Inhibition of this protein stops genomic stability such as to protect genomes from double-strand breaks or interstrand cross links [Kowalczykowski, 2000].

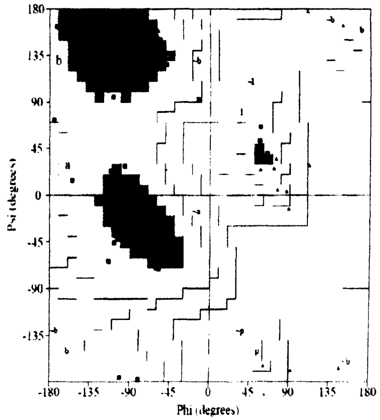
3.2 Homology Modeling: The 3D structure of DNA helicase RuvB protein of *C. pneumoniae* was modelled based on the crystal structure of *Thermotoga maritima* RuvB holliday junction branch migration motor protein[Putnam et al., 2001] (PDB ID: 1IN4) using Modeller9v2. The target protein sequence shows the following results when clustered based on a distance matrix with Modeller. (Figure 4)

```
Weighted pair-group average clustering based on a distance matrix:

      .---------------------------------------------------- 3cf2A
      |
      |                                             .--- 1hqcA
      |                                             |
      |          .------------------------------------- 1in4A
      |          |
      |          |        .---------------------------- 1iqpA
      |          |        |
      |          |        |    .----------------------- 1sxjC
      |          |        |    |
     .----------------------------------------------------- 1sxjD

     +----+----+----+----+----+----+----+----+----+----+----+----+
  100.7300  93.5075  86.2850  79.0625  71.8400  64.6175  57.3950
      97.1188  89.8962  82.6738  75.4512  68.2288  61.0062
```

Figure 4: Template proteins clustering based on a distance matrix.

The comparison in Figure 4 shows that 1hqc:A and 1in4:A are almost identical, both sequentially and structurally. However, 1in4:A has a better crystallographic resolution (3.2Å versus 1.6Å), eliminating 1hqc:A[yamada et al., 1997]. A second group of structures (1iqp:A, 1sxj:c) share some similarities. From this group, 1sxj [Bowman et al., 1997] has the poorest resolution leaving for consideration only 1iqp:A. 3cf2:A [Davies et al., 1997] is the most diverse structure of the whole set of possible templates. However, it is the one with the lowest sequence identity (26%) to the query sequence. Finally taken 1in4:A over instead of 1iqp:A[Oyama et al., 1997] because of its better resolution versus 1.6Å , its better crystallographic R-factor (23.4%) and higher overall sequence identity to the query sequence (53%). The protein was modelled with best template 1in4. The final structure of modelled protein is shown in Figure 5.

Figure 5: Predicted 3-D structure of Holliday junction DNA helicase RuvB protein. The modeled protein is validated with the SAVES server. The results of the PROCHECK analysis indicate that a relatively low percentage of residues have phi/psi angles in the disallowed ranges, the quality of Ramchandran plots is acceptable [Ramachandran et al., 1963]. The percentage of residues in the "core" region of modeled was found to be 94.2%. The stereochemical quality of the model was found to be satisfactory. The Ramachandran plot of the modeled protein is shown in Figure 6.



Plot statistics

Residues in most favoured regions
[A,B,L] 275 94.2%
Residues in additional allowed regions
[a,b,l,p] 15 5.1%
Residues in generously allowed regions
[~a,~b,~l,~p] 1 0.3%
Residues in disallowed regions
1 0.3%
---- ---------------------------------------

Number of non-glycine and non-proline residues 292 100.0%
Number of end-residues (excl. Gly and Pro) 2
Number of glycine residues (shown as triangles) 30
Number of proline residues
13

Total number of residues
337

M/c bond lengths: 99.5% within limits 0.5% highlighted
M/c bond angles: 93.2% within limits 6.8% highlighted
Planar groups: 100.0% within limits 0.0% highlighted

Figure 6: Ramachandran plot of Holliday junction DNA helicase RuvB protein generated by PROCHECK. The most favored regions are colored red, additional allowed, generously allowed and disallowed regions are indicated as yellow, light yellow and white fields, respectively.

Errat analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a residue [Colovos and Yeates, 1993]. Errat is showing an overall quality factor of 80.938. Verify_3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures [Eisenberg et al., 1997]. Verify_3D results show 92.90% of the residues had an averaged 3D-1D score larger than 0.2 and passed the test.

3.3 Active Site Prediction and Docking Study: Active sites of the target protein were predicted using the CASTp active site prediction tool. The feasible active sites predicted by the tool are shown in Table 2.

Table 2: The predicted active sites of Holliday junction DNA helicase RuvB protein.

S. No.	Cavity Area	Cavity Volume
1	941.2	1333.5
2	474	1115.5
3	251	383.5
4	188.8	143.8
5	119.2	108.5
6	137.9	126.2
7	174.4	150.7
8	64.8	59.6
9	68.9	50.9
10	81.1	61.2

4. Conclusion

Seven common non-human homologous essential genes were identified among *C. pneumoniae* strains. Out of the identified targets Holliday junction DNA helicase RuvB plays important role in DNA recombination and repair. Inhibition of this protein stops genomic stability such as to protect genomes from double-strand breakage or inter-strand cross-linkage [Tuteja et al., 2006; Frick, 2003]. The non homogeneity of the sequence with the human genome rules out the cross interference of the drugs designed specifically against it with the host metabolom. A homology model of Holliday junction DNA helicase RuvB was built by homology modeling. The model was validated using Ramachandran's plot, VERIFY3D, ERRAT and PROCHECK and WHAT_IF programs to arrive at a reliable and stable model for structure based drug design. Validation programs showed that the homology model scores are similar to crystal structure of the template.

Acknowledgement
We are also thankful to The Department of Biotechnology, Government of India for the funds to set up a bioinformatics infrastructure facility at our Department of Biotechnology and Medical Engineering, at National Institute of Technology, Rourkela where this work has been performed.

References

Binkowski, T.A., Naghibzadeh, S. and Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Research, 31, 3352-3355.

Bowman, G.D., ODonnell, M. and Kuriyan, J. (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature , 429, 724-730.

Cazzola, M.A., Matera, M.G.B. and Blasi, F. (2004). Macrolide and occult infection in asthma. Current Opinion in Pulmonary Medicine, 10, 7-14.

Colovos, C. and Yeates, T.O. (1993). Verification of protein structures: patterns of non-bonded atomic interactions. Protein Science, 2, 1511-1519.

Davies, J.M., Brunger, A.T. and Weis, W.I. (2008). Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. Structure, 16, 715-726.

Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S. and Bandyopadhyay, D. (2006). In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biology, 6, 0005.

Eisenberg, D., Luthy, R. and Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. Methods in Enzymology, 277, 396-404.

Frick, D.N. (2003). Helicases as antiviral drug targets. Drug News Perspect, 16, 355.

Galperin, M.Y. and Koonin, E.V. (1999). Searching for drug targets in microbial genomes. Current Opinion in Biotechnology, 10, 571-578.

Grayston, J.T., Campbell, L.A., Kuo, C.C., Mordhorst, C.H., Saikku, P., Thom, D.H. and Wang, S.P. (1990). A new respiratory tract pathogen: Chlamydiapneumoniae strain TWAR. Journal of Infectious Diseases, 161, 618-625.

Hahn, D.L. (1999). Chlamydia pneumoniae, asthma, and COPD: what is the evidence?. Annals of Allergy, Asthma and Immunology. 83, 271-292.

Judson, N. and Mekalanos, J.J. (2000). Transposon-based approaches to identify essential bacterial genes. Trends in Microbiology, 8, 521-526.

Kalman, S., Mitchell, W., Marathe, R., Lammel C, Fan, J., Hyman, R.W., Olinger, L, Grimwood, J., Davis, R.W. and Stephens, R.S. (1999). Comparative genomes of Chlamydophila pneumoniae and Chlamydophila trachomatis. Nature Genetics, 21, 385-389.

Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. Trends in Biochemical Sciences, 25 , 156–65.

Kroemer, R.T., Doughty, S.W., Robinson, A.J. and Richards, W.G. (1996). Prediction of the three-dimensional structure of human interleukin-7 by homology modeling. Protein Engineering, 9, 493–498.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993), PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26, 283-291.

Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46, 3-26.

Moir, D.T., Shaw, K.J. and Hare, R.S. (1999). GF: Genomics and antimicrobial drug discovery. Antimicrobial Agents and Chemotherapy, 43, 439-446.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. Proteins:Structure,Function,and Bioinformatics, 12, 345-364.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998). Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. Journal of Computational Chemistry, 19, 1639-1662.

Oyama, T., Ishino, Y., Cann, I.K.O., Ishino, S. and Morikawa, K. (2001) Atomic Structure of the Clamp Loader Small Subunit from Pyrococcus furiosus, Molecular Cell, 8,455-463.

Pontius, J., Richelle, J. and Wodak, S.J. (1996). Quality assessment of protein 3D structures using standard atomic volumes. Journal of Molecular Biology, 264, 121-136.

Putnam, C.D., Clancy, S.B., Tsuruta, H., Gonzalez, S., Wetmur, J.G. and Tainer, J.A. (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. Journal of Molecular Biology, 311, 297-310.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology, 7, 95-99.

Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Engineering, 12, 85-94.

Sakharkar, K.R., Sakharkar, M.K. and Chow, V.T.K. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biology, 4, 0028.

Sali, A. and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 234, 779-815.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24, 4876-4882.

Tuteja, R. and Pradhan, A. (2006). Unraveling the 'DEAD-box' helicases of Plasmodium falciparum. Gene, 376(1), 1-12.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. Journal of Molecular Graphics, 8, 52-56.

West, S.C. (1997). Processing of recombination intermediates by the RuvABC proteins. Annual Review of Genetics, 31, 213–244.

Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H. and Morikawa, K. (2001) Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8. Proceedings of the National Academy of Sciences of the United States of America, 98, 1442-1447.

Yang, S.I., Lickteig, R.L., Estes, R., Rundell, K., Walter, G and Mumby, M.C. (1991). Control of protein phosphatase 2A by simian virus 40 small-tantigen. Molecular and Cellular Biology, 11(4), 1988-1995.

Zhang, R., Ou, H.Y. and Zhang, C.T. (2004). DEG: A Database of Essential Genes. Nucleic Acids Research, 32, D271-D272.

STATISTICAL METHODS IN BIOINFORMATICS RESEARCH WITH SPECIAL EMPASIS ON GENOMIC AND MICROARRAY DATA ANALYSIS OF FISHES

A.K Roy[1]

1. Introduction

Bioinformatics is a new scientific discipline that combines biology, computer science, mathematics, and statistics into a broad-based field that will have profound impacts on all fields of biology. Bioinformatics is expected to substantially impact on scientific, engineering and economic development of the world. Research and development in bioinformatics and computational biology require the cooperation of specialists from the fields of biology, computer science, mathematics, statistics, physics, and such related sciences. It is the comprehensive application of mathematics (e.g., probability and graph theory), statistics, science (e.g., biochemistry), and computer science (e.g., computer algorithms and machine learning) to the understanding of living systems. Bioinformatics is fast emerging as an important discipline for academic research and industrial application. The large size of biological datasets, the inherent complexity of biological problems and the ability to deal with error-prone data all result in special requirements, such as large memory space and huge computation time.

In the post genomic age with the advent of huge amount of Genomic data we are now entering the and the focus in Genomic research is switching from sequencing to using the genome sequences in order to understand their construct and functioning. The management of these data have become very cumbersome and requires statistical methods and knowledge visualization tools and techniques for proper understanding of the hidden facts. **Visualization** has been used for over a thousand years for creating images, diagrams and animations to communicate important information. Knowledge visualization aims at transfer of facts, insights, experiences, attitudes, values, expectations, perspectives, opinions, and predictions by using various visualization techniques or visual representation of useful information like table, matrix, charts, plots, graphs, maps, etc. There are evidences of using various multivariate statistical methods like multidimensional scaling, correspondence analysis, principle component analysis, clustering methods etc. for analysis of multivariate data and genomic data are multivariate in nature. In this paper it has been attempted to represent an overview on analysis of codon usage variation and micro array data analysis.

2. Genomic Sequence Analysis

Multivariate analysis methods such as Correspondence Analysis and Principle Component Analysis have often been used to identify major trends of variation in synonymous codon usage among inter or intra specific genes. Genetic codes are degenerate meaning most of amino acids can be encoded by more than one

[1] **Correspondance**: A. K. Roy; Ex. Coordinator, BTIS & Head Social Science Section; Central Institute of Freshwater Aquaculture; Kausalyaganga, Bhubaneswar-751002, India
E-mail: akroy1946@yahoo.co.in

codon (triplet of nucleotides); such codons are synonymous and usually differ by one nucleotide in the third position. The alternative synonymous codons are not used with similar frequency and their usage varies among different genes. Hence there is the presence of Relative Synonymous Codon Usage values.

2.1 Sequence Data

The gene sequences downloaded from Genbank of National Center for Biotechnology Information that is the nucleotide sequence repository known worldwide, serves as the secondary data used for analysis. Relative synonymous codon usage value of each gene is extracted separately using the Codon Usage software of Sequence Manipulation Suite. These are the raw data that needs to be normalized to be used before application of various multivariate analyses.

2.2 Relative Synonymous Codon Usage

The first stage normalized data are absolute codon frequencies to avoid biases, such as gene length and amino acid usage, which may mask variation in synonymous codon usage. The effect of gene length can be avoided by dividing the frequencies of codons in a gene by the total codon frequency in the same gene, but such normalized data can be affected by amino acid usage in the sequence. To avoid the effect of amino acid usage, the codon frequencies are often normalized for each individual amino acid. However, two such sets of frequently used normalized data, relative codon frequencies (the number of times the codon is observed i.e. the frequency of occurrence of the codon in a gene) and relative synonymous codon usage. RSCU values are a reflection of how often a particular codon is used relative to the expected number of times that codon would be used in the absence of codon usage bias. It is the relative frequencies of synonymous codons to code for same amino acid. Synonymous codons form the subset of codons coding for a single amino acid.

2.3 Correspondence Analysis of Sequence Data

Correspondence Analysis is a multivariate exploratory graphical technique used to describe the relationship between the row and column variables of a contingency table. A matrix is constructed keeping genes as row variable and codons as column variable and RSCU values as weight variables that is uploaded into the statistical package SPSS to carry out Correspondence Analysis.

2.3.1. Construction of Matrix

Table 1. Hypothetical table with genes forming rows and codon representing columns

Codons	Codan1	Codan2	Codan3	...	Codan62	Codan63	Codan64
Genes							
Gene 1	$RSCUg_1c_1$	$RSCUg_1c_2$	$RSCUg_1c_3$		$RSCUg_1c_{62}$	$RSCUg_1c_{63}$	$RSCUg_1c_{64}$
Gene 2	$RSCUg_2c_1$	$RSCUg_2c_2$	$RSCUg_2c_3$		$RSCUg_2c_{62}$	$RSCUg_2c_{63}$	$RSCUg_2c_{64}$
Gene 3	$RSCUg_3c_1$	$RSCUg_3c_2$	$RSCUg_3c_3$		$RSCUg_3c_{62}$	$RSCUg_3c_{63}$	$RSCUg_3c_{64}$
. . .							
Gene 98	$RSCUg_{98}c_1$	$RSCUg_{98}c_2$	$RSCUg_{98}c_3$		$RSCUg_{98}c_{62}$	$RSCUg_{98}c_{63}$	$RSCUg_{98}c_{64}$
Gene 99	$RSCUg_{99}c_1$	$RSCUg_{99}c_2$	$RSCUg_{99}c_3$		$RSCUg_{99}c_{62}$	$RSCUg_{99}c_{63}$	$RSCUg_{99}c_{64}$
Gene 100	$RSCUg_{100}c_1$	$RSCUg_{100}c_2$	$RSCUg_{100}c_3$		$RSCUg_{100}c_{62}$	$RSCUg_{100}c_{63}$	$RSCUg_{100}c_{64}$

The result generated in form of many tables and plots. It gives output in form of correspondence table, table of dimensionality, profiles and distances, row and column scores, eigen values, mass, centroid, singular values, confidence statistics, plots of symmetrical normalization. A correspondence table is any two-way table whose cells contain some measurement of correspondence between the rows and the columns. The measure of correspondence can be any indication of the similarity, affinity, confusion, association, or interaction between the row and column variables.

A correspondence table is a cross tabulation, where the cells contain frequency counts. Correspondence analysis computes the row and column profiles to determine the distance between categories. Row profiles indicate the proportion of the row category in each column category. It corresponds to the cell contents divided by their corresponding row total. Similarly column profiles indicate the proportion of the column in each row category. Likewise, they are the elements divided by the column marginal. Mass is a measure that indicates the influence of an object based on its marginal frequency. Point with lagre mass pull the centroid strongly to their location. Points with small mass pull the row centroid only slightly to its location. Correspondence analysis solution represents the relationship between row and column variables in as few dimensions as possible. It is frequently useful to look at the maximum number of dimensions to see the relative contribution of each dimension. . Eigen values (inertia) are the percent of variance explained by each dimension. The row and column scores are the coordinates of the row and column points in the correspondence map.

2.3.2. Correspondence Map

Correspondence Analysis generates a variety of plots that graphically illustrate the underlying relationships between categories and between variables. Biplot is the scatter plot of the row and column scores for the two-dimensional solution.

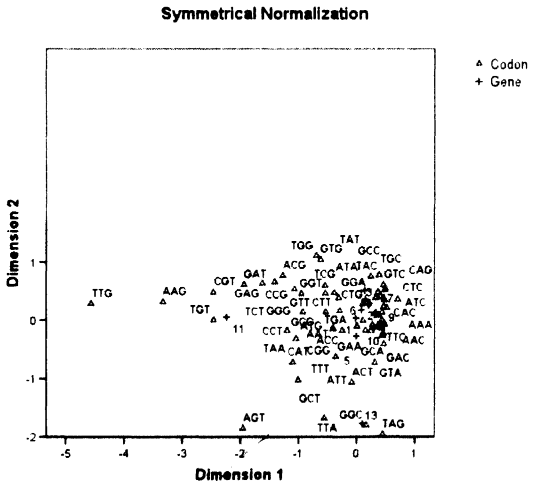


Fig. 1Biplot between genes and codons of C. carpio generated by Correspondence Analysis of SPSS

In the example given in this paper, a plot between 13 genes and 64 codons of Cyprinus carpio is displayed. The plot reveals that except two genes that are located far away from the origin, all other genes follow a common trend. The genes 11 is associated with codons TGT, AAG, CGT and gene 13 is associated with codons GGC, TTA, TAG while all other genes viz. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 12 prefer GC**G**, CG**G**, TG**C**, TA**C**, ATA, TC**G**, GGA, GT**C**, CA**G**, CT**C**, CT**G**, CTT, AT**C**, CA**C**, TGA, AT**G**, AAT, AC**C**, GAA, TT**C**, AA**C**, GCA, GA**C**, ACT, GTA etc over other codons for synthesizing their respective protein. It is clear from this that the genes are baised to GC3 content.

2.4. Principal Component Analysis of Sequence Data

Principal components analysis can be thought of as a method of dimension reduction. A set of variables is analyzed to reveal major dimensions of variation. The original data set can then be replaced by a new, smaller data set with minimal loss of information. The method reveals relationships among variables, among cases, and among variables and cases. The criterion used by categorical principal components analysis for quantifying the observed data is that the object scores (component scores) should have large correlations with each of the quantified variables.

Joint Plot of Category Points of Genes and Codons

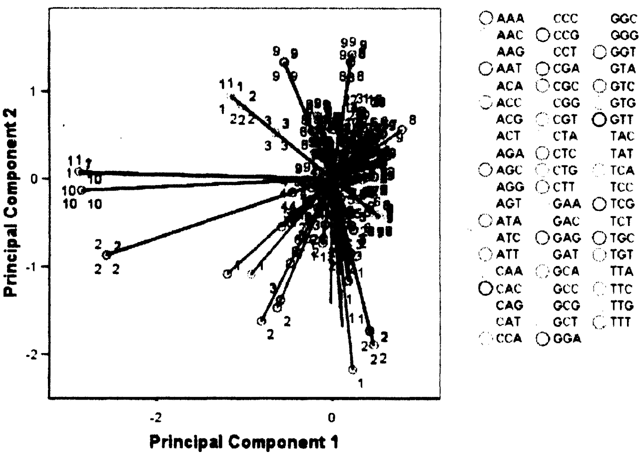


Fig. 2 An example of Principal Component Analysis generated by SPSS.

Glancing at the quantifications in the joint plot of the category points, you can see that some of the categories of some variables were not clearly separated by the categorical principal components analysis as cleanly as would have been expected if the level had been truly ordinal. The joint plot of category points resembles the plot for the component loadings, but it also shows where the endpoints are located that correspond to the lowest quantifications. Figure 2 displays an example of categorical principle component analysis carried out between genes and codons using codons as ordinal variables.

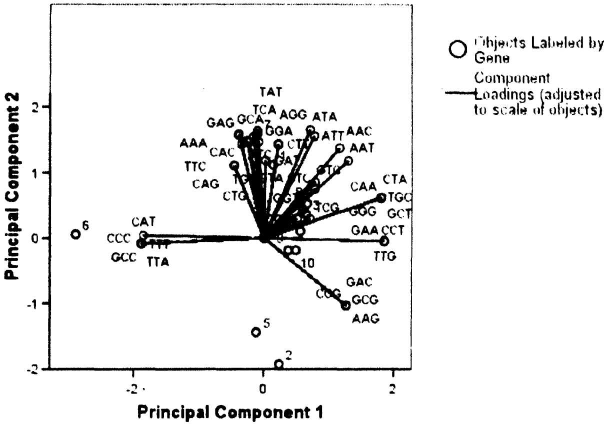


Fig. 3 An example of Principal Component Analysis generated by SPSS

3. MICROARRAY ANALYSIS

Microarray provides insight into the transcriptional state of the cell, measuring RNA levels for thousands of genes at once. It requires preparation of mRNA from cells growing under certain experimental conditions. Subsequently, the labeled cDNA mixture is hybridized to the microarray. After detection of the signals, image analysis programs are used to determine spot intensities. Data produced may be regarded as a table, each row representing a gene, each column standing for an experimental condition. Multiple measurements for each condition, involving repeated sampling, labeling, and hybridization, offer the opportunity of extracting more robust signals. The columns of the table have to undergo a normalization procedure, correcting for affine-linear transformation among the columns. all of the methods recently used for microarray data analysis would result in an outline of applied statistics, however. Most methods fall into one of three groups, namely clustering, classification, and projection methods. Examples of clustering techniques are k-means clustering, hierarchical clustering, and self-organizing maps. Classification methods take as input a grouping of objects and aim at delineating characteristic features common and discriminative to the objects in the groups. Examples of classification methods range from linear discriminant analysis to support vector machines or classification and regression trees.

3.1. Generation of Microarray Images

There are a variety of microarray platforms that have been developed to accomplish and the basic idea for each is a glass slide or membrane that is spotted or "arrayed" with DNA fragments or oligonucleotides that represent specific gene coding regions. Purified RNA is then fluorescently- or radioactively labeled and hybridized to the slide/membrane. In some cases, hybridization is done simultaneously with reference RNA to facilitate comparison of data across multiple experiments. After thorough washing, the raw data is obtained by laser scanning or autoradiographic imaging. The final step of the laboratory process is to produce an image of the surface of the hybridized array. Fluorescently labeled microarrays can then be "read" with commercially available scanners scanning confocal microscopes with lasers exciting at wavelengths specifically for Cy3 and

Cy5. The amount of signal emitted is directly in proportion to the amount of dye at the spot on the microarray and these values are obtained and quatitated on the scanner. This results in the image that is typical microarray picture.

3.2. Preprocessing

Preprocessing of microarray data comprises analytical or transformational procedures that is directed at resolving the systematic error and bias.

3.2.1. Checking the Background Information: Background and foreground intensities together form the spot intensity. If the spot intensities are dependent on the background intensities, it is possible either not to apply any background correction to the data or discard the deviating observations from further analyses. The background corrected intensity values are calculated spotwise by subtracting the background intensity from the spot (foreground) intensity.

3.2.2. Calculation of Expression Change: There are three commonly used measures of expression change. Intensity ratio is the raw expression value, and log ratio and fold change are transformationally derived from it.

(i) Intensity Ratio: After background correction, expression change is calculated. The simplest approach
is to divide the intensity of a gene in the sample by the intensity level of the same gene in the control. Intensity ratio can be calculated from background corrected
or uncorrected data (here we use corrected data) The formula for two-color data is

$$\text{Intensity ratio} = \frac{Cy3'}{Cy5'}$$

For affymetrix data, substitute Cy3' and Cy5' with the appropriate intensitiesfrom the sample and control chips. This intensity ratio is one for an unchanged expression, less than one for down regulated genes and larger than one for up-regulated genes. The problem with intensity ratio is that its distribution is highly asymmetric or skewed. Up-regulated genes can take any values from one to infinity (1 to ∞) whereas all down regulated genes are squeezed between zero and one (0 to 1). Such distributions are not very useful for statistical testing.

(ii) Log Ratio: To make the distribution more symmetric (normal-like) log-transformation can be applied. The most commonly used log-transformation is 2-based (*log* 2), but it does not matter whether you use natural logarithm (*loge*) or base 10 logarithm (*log*10) as long as it is applied consistently to all the samples. The formula for the calculation of *log*2-transformation is:

$$\text{Log ratio} = log_2(\text{Intensity ratio}) = log_2\left(\frac{Cy3'}{Cy5'}\right)$$

After the log-transformation, unchanged expression is zero, and both upregulated and down-regulated genes can take values from zero to infinity. Log ratio has some nice properties compared with other measures of expression. It makes skewed distributions more symmetrical, so that the picture of variation becomes more realistic. In addition, normalization procedures become additive.

For example, we have the following intensity ratio results for two replicates (A and B) after normalization:

$$GeneA = \frac{120}{60} = 2.0$$

$$GeneB = \frac{30}{60} = 0.5$$

The mean of these replicates is 1.25 instead of 1, which would have been expected. If the 2-based logarithmic transformation is applied, the log ratios are:

$$GeneA = log_2\left(\frac{120}{60}\right) = 1.0$$

$$GeneB = log\left(\frac{30}{60}\right) = -1.0$$

The mean of these log ratios is 0, which corresponds to the mean intensity ratio of 1. Although log-transformation is not always the best choice for microarray data, it is used because the other transformations lack this handy additive property. One downside of the log-transformation is that it introduces systematic errors in the lower end of the expression value distribution.

3.2.3. Replicates: If the experiment includes replicates, their quality can be checked with simple methods, using scatter plots and pairwise correlations or hierarchical clustering techniques.When the distribution of the intensity values is skewed, the median characterizes the central tendency better than the mean. The median and mean can also be used to check the skewness of the distribution. For symmetrical distributions mean and median are approximately equal.

3.2.4.Outliers and Filtering bad & Uninteresting Data: Outliers in chip experiments can occur at several levels. There can be entire chips, which deviate from all the other replicates. Or there can be an individual gene, which deviates from the other replicates of the same gene. Outliers should consist mainly of quantification errors. In practise, it is often not very easy to distinguish quantification errors from true data, especially if there are no replicate measurements. If the expression ratio is very low (quantification errors) or very high (spot intensity saturation), the result can be assumed to be an artifact, and should be removed. Most of the actual outliers should be removed at the filtering step (those that have too low intensity values). This is often equivalent to a filtering, where observations with too low or high intensity values are excluded from further analyses.

3.2.5. Linearity: Linearity means that in the scatter plot of channel 1 (red colour) versus channel 2 (green colour), the relationship between the channels is linear. It is often more informative to produce a scatter plot of the log-transformed intensities, because then the lowest intensities are better represented in the plot. In this kind of a plot, the data points fit a straight line, if the data is linear.Another way to test linearity is to plot the log ratio versus the intensity of one channel. This should be done for both channels independently.

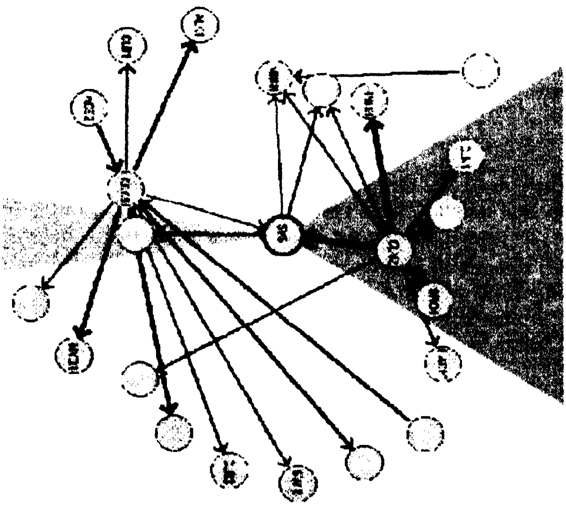
3.3. Statistical Analysis of Microarray data

The basic concepts of clustering are to try to identify and group together similarly expressed genes and then try to correlate the observations to biology. The idea is that co-regulated and functionally related genes are grouped into clusters. Clustering provides the framework for this analysis. Principles of clustering:

1. Clustering organizes the data into a small number of homogeneous groups. The methods are used to find similar expression motifs irrespective of the expression level. Therefore, both low and high expression level genes can end up in the same cluster if the expression profiles are correlated by shape.
2.Clustering methods can be grouped as supervised and unsupervised. Supervised methods assign some predefined classes to a data set, whereas in unsupervised methods no prior assumptions are applied.

Different types of clustering involves Hierarchical clustering, K-means, self-organizing maps (SOMs), principal component analysis (PCA), and K-nearest neighbor (KNN) have been commonly used. There are also other methods, such as Correspondence Analysis, Fuzzy K-Means Clustering Multidimensional scaling (MDS), minimum description length (MDS), gene shaving (GS), decision trees, and support vector machines (SVMs).

Analysis of multiple arrays: Networks



From Friedman et al (2000). *Using Bayesian Networks to Analyze Expression Data*, Journal of Computational Biology, 7:601-620.

3.3.1. Hierarchical Clustering

The hierarchical clustering can be represented as a tree, or a dendrogram. Branch lengths represent the degree of similarity between the genes. Hierarchical clustering consists of two separate phases. The hierarchical clustering is performed as follows; Each gene is assigned to a cluster of its own. The closest pair of clusters is found and merged into a single cluster. The distances (similarities) between the new clusters are computed and each of the old clusters using either single, average or complete linkage method. Above steps are repeated until all genes are clustered.

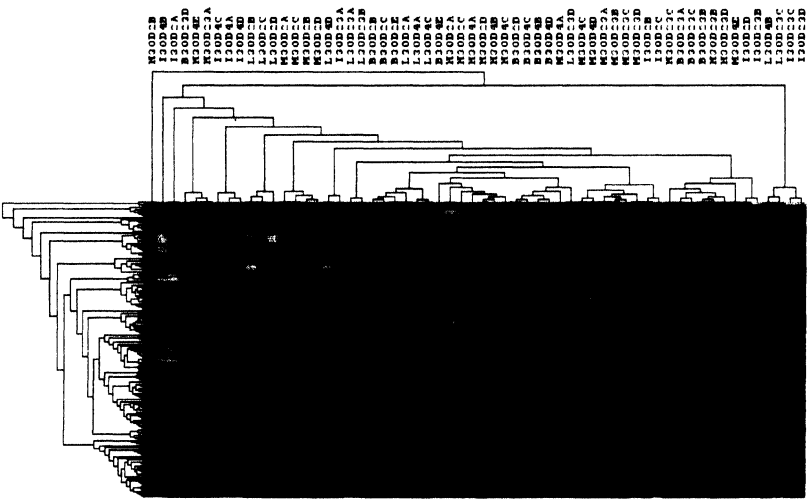


Fig4. An Example of Hierarchical clustering of C. carpio obtained using MeV software

In single linkage clustering, the distance between one cluster and another is considered to be equal to the shortest distance from any member of one cluster to any member of the other cluster. In complete linkage clustering, the distance between one cluster and another cluster is considered to be equal to the longest distance from any member of one cluster to any member of the other cluster. In average linkage, the distance between one cluster and another cluster is considered to be equal to the average distance from any member of one cluster to any member of the other cluster.

3.3.2. Self-Organizing Map

Kohonen's self-organizing map (SOM) is a neural net that uses unsupervised learning. SOM tries to learn to map similar input vectors (gene expression profiles) to similar regions of the output array of nodes. The method maps the multidimensional distances of the feature space to two-dimensional distances in the output map. In SOMs, the number of clusters has to be predetermined. The dimensions of the two-dimensional grid or array give the value. One has to experiment with the actual number of clusters and find such array dimensions that there is a minimum number of poorly fitting genes in the clusters. The actual number of clusters is also difficult to assign, but the square root of the number of genes is a good initial estimate.

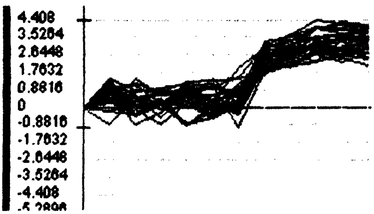


Fig. 5 An example of self-organizing map (Source; Hovatta et. al.)

3.3.3. K-means Clustering

K-means is a least-squares partitioning method for which the number of groups, K, has to be provided. The algorithm computes cluster centroids and assigns each object to the nearest centroid. However, it is also possible to estimate K from the data, taking the approach of a mixture density estimation problem. The genes are arbitrarily divided into K centroids. The reference vector, *i.e.* location of each centroid, is calculated. Each gene is examined and assigned to one of the clusters depending on the minimum distance. The centroid's position is recalculated. Above steps are repeated until all the genes are grouped into the final required number of clusters.

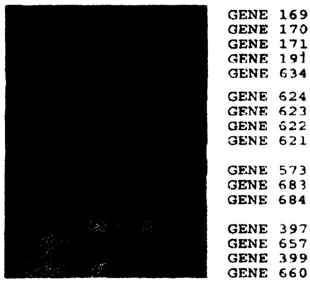


Fig. 6 An example of k-means clustering with four clusters (Source; Hovatta et. al.)

K-means partitioning is a so-called NP-hard problem, thus there is no guarantee that the absolute minimum of the objective function has been reached. Therefore, it is good practice to repeat the analysis several times using randomly selected initial group centroids, and check whether these analyses produce comparable results.

3.3.4. Principal Component Analysis of Microarray

Objectives of principal component analysis (PCA) are to discover or to reduce the dimensionality of the data set and to identify new meaningful underlying variables. PCA transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The basic idea in PCA is to find the components that explain the maximum amount of variance possible by n linearly transformed components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA can be also applied when other information in addition to the actual expression levels is available (this applies to SOM and K-means methods as well).

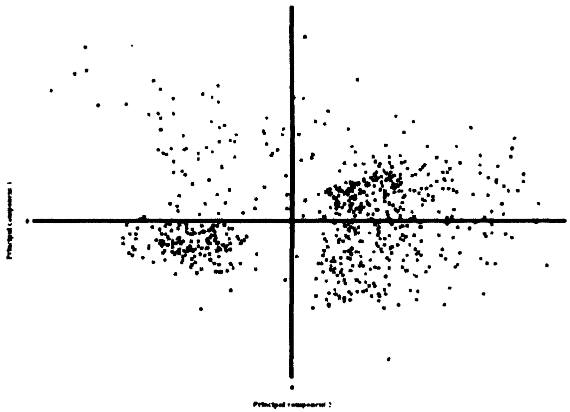


Fig. 7 An example of principal component analysis. The two most significant principal components have been selected as the axes of the plot (Source; Hovatta et. al.)

3.3.5. Correspondence Analysis

Correspondence analysis is an explorative method to study associations between variables. It directly visualizes associations between genes and hybridizations. Unlike many other methods, CA does not require any prior choice of parameters. Like principal components, it displays a low-dimensional projection of the data. However, in this case, both genes and samples can be projected onto the same space, revealing associations between them. Correspondence analysis requires an expression matrix with no missing values. Therefore, any missing values have to be imputed first. We use the k-nearest neighbors algorithm to impute missing values. The only user input in the initialization dialog is the desired number of neighbors for imputation. Genes that lie close to one another on the plot tend to have similar profiles, regardless of their absolute value. The same is true for samples. If some genes and samples lie close to one another on the plot, then these genes are likely to have a high expression in the nearby samples relative to other samples that are far away on the plot. On the other hand, if a set of genes are on the opposite side of the plot from a set of samples relative to the origin, then the expression of that set of genes is likely to be depressed in those samples relative to samples that might be positioned close to those genes. The farther the points are from the origin, the stronger the association between genes and samples. Correspondence analysis works by decomposing a matrix of chi-squared values derived from the rows and columns of the expression matrix. The first two or three axes are the most informative in showing associations among genes and experiments. The amount of information explained by a given axis is quantified by its inertia, which may be thought of as the proportion of the total chi-squared value of the matrix explained by that axis. The inertia values are provided under the corresponding node under the main COA analysis node.

4. Conclusion

Demonstration of statistical methods with illustrated examples will help future researchers in carrying out complex statistical analysis using statistical packages and bioinformatics tools and techniques to visualize fact through data mining and exploratory data analysis.

References

Ames, B., and Hartmann, P. (1963). The histidine operon. Cold Spring Harbor Symposium Quantitative Biology, **28:** 349-356.

Bonnet, E; Fostier, A; Bobe, J*. (2007) Microarray-based analysis of fish egg quality after natural or controlled ovulation. BMC Genomics [BMC Genomics]. **8(55):** 17.

Clark, MD; Hennig, S; Herwig, R; Clifton, SW; Marra, MA; Lehrach, H; Johnson, SL. (2001). An Oligonucleotide Fingerprint Normalized and Expressed Sequence Tag Characterized Zebrafish cDNA Library. Genome Research [Genome Res.], **11(9):**1594-1602.

Derome, N; Duchesne, P; Bernatchez, L*. (2006). Parallelism in gene transcription among sympatric lake whitefish (Coregonus clupeaformis Mitchill) ecotypes Mol. Ecol.,**15 (5):**1239-1249.

Greenacre,M.J. (1984). Theory and Applications of Correspondence Analysis. Academic Press, London.

Holm, L. (1986). Codon usage and gene expression. Nucleic Acid Res., **14:** 3075- 3087.

Lafay, B., Lioyd, A.T.,Mc Lean, M.J., Devine,K.M.,Sharp,P.M. and Wolfe,K.H. (1999). Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutataional biases. Nucleic Acids Res., **27:** 1642-1649.

Larkin, Patrick; Villeneuve, Daniel L; Knoebl, Iris; Miracle, Ann L; Carter, Barbara J; Liu, Li; Denslow, Nancy D; Ankley, Gerald T. (2007) Development and Validation of a 2,000-Gene Microarray for the Fathead Minnow (Pimephales Promelas). Environ. Toxicol. Chem. **26(7):**1497-1506.

McLean, TI; Crawford, DL. (2006). Towards a Genetic Understanding of Karenia brevis: Using Microarrays to Discover Gene Expression Patterns as They Relate to K. brevis Ecology EOS, Transactions, American Geophysical Union [EOS Trans. Am. Geophys. Union]. **87(36):** suppl., [np]. suppl.

McInerney, J.O. (1998). GCUA: General Codon Usage Analysis. Bioinformatics, **14:**372-373.

Medigue, C., T. Rouxel, P. Vigier, A. Henaut and Danchin, A. (1991). Evidence for horizontal gene transfer in *Escherichia coli* speciation. Journal of Molecular Biology, **222:** 851-856.

Morrison, D.A., Ellis, J., and Johnson, A.M. (1994) An empirical comparison of distance matrix techniques for estimating codon usage divergence. Journal of Molecular Evolution **39:** 533-536.

Moszer, I., Rocha, E.P.C. and Danchin, A. (1999). Codon usage and lateral gene transfer in *Bacillus subtilis* genome. Microbiology, **141:** 261-268.

Osawa, S. and Jukes, T. H. (1989). Codon reassignment (codon capture) in evolution. Journal of Molecular Evolution, **29:** 271-278.

Quentin, Y. (1989). Successive waves of fixation of BI variants in rodent lineage history.J.Mol.Evol., **28:** 299-305.

Romero, H., Zavala, A. and Musto, H. (2000). Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acid Res., **28:** 1084-2090.

Roy, A. K, Martha, S. R and Rath, S. N. (2007) Correspondence Analysis of Codon Usage in *Labeo rohita*. International Symposium on Chromosome to Genomes (c2g) at CCMB, Hyderabad, (Abs. No. 9)

Thioulouse, J., Chessel, D. and Dolédec, S. (1997). ADE-4: A multivariate analysis and graphical display software. Stat. Comput., **7:** 75–83.

Tsoi, SCM; Cale, JM; Bird, IM; Ewart, V; Brown, LL; Douglas, S. (2003). Use of Human cDNA Microarrays for Identification of Differentially Expressed Genes in Atlantic Salmon Liver During Aeromonas salmonicida Infection. Marine Biotechnology [Mar. Biotechnol.], **5 (6):** 545-554.

Weidinger, G; Thorpe, CJ; Wuennenberg-Stapleton, K; Ngai, J; Moon, RT*.(2005). The Sp1-Related Transcription Factors sp5 and sp5-like Act Downstream of Wnt/ beta -Catenin Signaling in Mesoderm and Neuroectoderm Patterning Current Biology [Curr. Biol.], **15(6):** 489-500.

ROLE OF SOCIOECONOMIC AND FARM SPECIFIC VARIABLES IN AQUACULTUE- AN ECONOMETRIC APPROACH

A. K. Roy[1]

1. 1 Introduction

India is endowed with freshwater aquaculture resources of 2.35 million hectare of ponds and tanks, 1.3 million hectare of oxbow lakes and derelict waters, 0.19 million km of rivers and canals, 3.15 million ha of reservoirs (Anon, 2006 a). Ponds and tanks are the prime resources for aquaculture, however only about 0.8 - 0.9 million ha is used for aquaculture currently. Indian aquaculture has demonstrated a six and half fold growth over the last two decades with freshwater aquaculture contributing over 95% of the total aquaculture production. Three Indian major carps namely catla (*Catla catla*), rohu (*Labeo rohita*) and mrigal (*C. mrigala*) contribute % the bulk of production of over 1.8 million tones (FAO, 2005). The technologies of induced carp breeding and polyculture in static ponds and tanks virtually revolutionized the freshwater aquaculture sector raising the average Indian production from still water ponds from 600 kg/ha/yr in 1970 to over 2200 kg/ha/yr at present and turned the sector into a fast growing industry. Research and development activities by Government of India through establishment of Freshwater Aquaculture Research Institute and network of Fish Farmers Development Agencies have been the principal vehicles for this development besides additional support by several other organizations, departments and financial institutions. Latest development in this sector is the establishment of National Fisheries Development Board with huge budget provision to improve production, processing, storage, distribution and marketing of fisheries products (Anon, 2006 b). The total aquaculture production of 2.47 million tones that constitute 4.2% of global production was valued at US$ 2.9 billion tones (FAO, 2006). If aquaculture is to play a vital role in ensuring future fish availability for food security and nutrition in India, this sector has to develop and grow in an economically viable and environmentally sustainable fashion. That is possible through improved water management, better feeding strategies, genetic improvement of cultivable species and better health management leading to enhanced productive efficiency of aquaculture at farm level. Therefore measuring Technical Efficiency (TE) at farm level identifying factors associated with efficient production and assessing potential for and sources of future improvements are need-of-the-hour for establishing sustainable aquaculture. Instead of increasing the use of inputs to increase production, efforts should be made towards output growth through improved TE, which means to produce more by utilizing inputs at hand more efficiently.

Keeping these points in view, an attempt was made to estimate the technical efficiency of aquaculture farms of Kolleru Lake area, which is known as the Carp Pocket of India, producing an estimated 70,000 tonnes per year (Anjaneyulu, Y.,

---

[1] **Correspondence:** Ex. Coordinator, BTIS & Head Social Science Section; Central Institute of Freshwater Aquaculture Kausalyaganga, Bhubaneswar-751002, India
E-mail : akroy1946@yahoo.co.in

2003). Kolleru Lake (16° 32′ and 16° 47′N latitude and 81° 5′ and 81° 21′E longitude) is the largest freshwater wetland ecosystem located between the major perennial rivers Godavari and Krishna in South India. It is 955 sq. km at 3.8 m msl contour level. Water depth in the lake varies between 1 and 2.5 m for most of the year and reaches a maximum depth of 3 to 4 meters during high floods and is reported to very productive for aquaculture with favourable temperature ranging between 18-49°C. There are 122 villages in the lake area having a population of nearly two lakh and the majority of them all fishermen community. Primary occupation of the villagers is aquaculture and fishing. Reports are available on lake flora and fauna, integrated development, ecodevelopment, fisherman population, fisheries, eutrophication and disease (Chacko *et.al* 1952; Murthy 1973; Dutt, 1976; Seshavatharam and Dutt, 1978; Ramakrishnan, 1980; INCOR, 1983; Reddy and Suryanarayan, 1983; Rama Rao, 1986-88; Ghosh *et al.*, 1990; Gopal Krishnayya and Lal, 1990). Further, documents covering research on physico-chemical properties, limnological aspects, biodiversity, geography, ecology, capture and culture fisheries, environmental qualities, status of freshwater aquaculture, strategic management plans, heavy metal concentrations in fishes, pesticide and insecticide pollution; government's crackdown order breaking fish tanks (Anjaneyulu, 2003; Padmavathi and Prasad, 2003; Rama Murthy and Swamy, 2003; Vidyanath, 2003; Rao, 2003; Khasim and Rao, 2003; Rao and Padmaja, 2003; Ramani and Anjaneyulu, 2003; Rao and Reddy, 2003; Rao, 2003; Sarangi *et al.*, 2004 and Madhavi, 2005) are also available. A report highlighting the technological changes of Kolleru Lake area covering pond construction and renovation, weed control, application of fertilizers, species combination, stocking density, use of supplementary feed, biomass and disease checking and control, harvesting, live fish transport, marketing and market spread, impact of technological changes on employment generation, growth of ancillary industries, environmental impacts, education, operational economics is available (Anon, 2002). Scientific studies on statistical and econometric aspects like development of sampling frame, statistics of carp culture, socio economic aspects, economic performance of aquaculture, factor share, database on socioeconomic aspect, farm size aquaculture productivity relationship, break even analysis, marketing channel are also available (Roy *et al.*, 2002; 2001 a; 2001 b; 2002 b; 2004 a; 2004 b; Sahoo *et al.*, 2000; 2001 a; Saradhi *et al.* 2002; Roy *et al.*, 2006).

In recent years, Farmers of Kolleru Lake area have intensified their effort towards aquaculture and as a result demand for natural resources like water, land and seed have increased. This phenomenon might have adverse environmental effects which may affect sustainability. Under that situation, enhanced production through improvement of technical efficiency may lead to sustainable production and consistent supply, through efficient use of resources and productivity at farm level. Therefore, it is necessary to measure technical efficiency (TE) at farm level, to identify important factors responsible for efficient production process. TE may be estimated either through parametric (frontier analysis) and nonparametric methods (DEA). A review of literature reveals that estimation of technical efficiency applying frontier analysis is very limited in aquaculture sector in India (Jayaraman, 1998) and abroad (Sharma and Leung, 1998; Sharma and Leung, 2000 a, b; Dey *et al.* 2000; Iinuma *et al.* 1999). The present study is aimed at measuring the technical efficiency and also to estimate the impact of socio-economic and farm specific variables on technical efficiency of carp farms of Kolleru Lake area.

2. Data and variables

2.1. Data sources

The present analysis was carried out on a data set of 221 aquaculture farms that was collected using multistage stratified random sampling covering 73 revenue villages spread over 9 Mandals of West Godavari and Krishna districts of Andhra Pradesh under the AP-Cess fund project entitled "Technological innovations in aquaculture and its effects on sustainability of farming systems in Andhra Pradesh" funded by Indian Council of Agricultural Research (ICAR) during 1998-2002. Farmers of Kolleru Lake generally practice semi-intensive composite culture of two Indian major carps namely catla (*Catla catla*) and rohu (*Labeo rohita*) at a species ratio of 1:4 and at a stocking density ranging from 5 - 10 thousand advanced fingerling/ha with higher level of input recommended for composite fish culture adopting single stocking and single harvesting techniques. The earlier practice of five species poly culture of carps (IMC and exotic) is now replaced by a two species (Rohu and Catla) culture system (Padmavathi and Prasad 2003). This technology is the modified version of composite fish culture technology (Chaudhuri *et al.*1975) that the farmers of Kolleru Lake area have perfected with respect to species composition and input use, which were driven by growth of cultivable species, market demand, consumer preference, keeping quality and economics. The technology delivers a yield of 7-15 tones/ha/yr.

2.2. Input and output variables

All input and output variables used in the stochastic frontier production function model in this study are clearly defined along with measurement units (Table 1). Summary statistics of organic manure, inorganic fertilizer, lime, chemicals and drugs, feed applied, types of labor engaged and socio-economic profiles like religion, caste, age, no. of children, education, experience, asset, etc. of the farm operators are furnished in table-2.

2.3. Sample Characteristics

The average pond size of our sample was 8.7 ha and the range was 0.8 - 61 ha. The primary occupation of 79% of the farmers is aquaculture. 96% of the farmers receive water from the adjoining irrigation canals and 4% take water from natural sources. 92% of the farmers are facility owners or survey revealed that 83% farmers reported renovating now often their ponds as part of pond management practices. All farmers cultured catla and rohu at a species ratio of 1:4 and at an average stocking biomass of 1939 kg/ha. Single stocking and single harvesting practices per year was resorted to by 95% of the farmers. Feed 97% of farmers applies was using perforated bags hanged from poles and submerged. Average yield of fish was 10.63 t/ha/year.

3. Empirical Model for Estimation of Technical Efficiency and Test of Hypothesis

It is mentioned that the level of technical efficiency of a particular farm is characterized by the relationship between observed production and some ideal production (Greene, 1993). The frontier production function defines potential output that can be produced by a farm with a given level of input and technology. Farrell (1957) first developed frontier production models in aquaculture. Since then several approaches to efficiency assessment have been developed that can be classified into two broad categories: parametric and nonparametric. Frontier techniques have been widely used in determining the productive performance of farms which may be estimated either by a stochastic frontier production function (Aigner *et al.* 1977; Meeusen and van der Broeck 1977) or non-parametric linear programming approach, known as data envelopment analysis (DEA). The stochastic frontier approach is preferred for assessing efficiency in agriculture because of the inherent stochasticity involved (Kirkley *et al.* 1995; Coelli *et al.* 1998). There are several reviews of the applications of the stochastic frontier approach in agriculture (Battese and Coelli 1992; Bravo-Ureta and Pinheiro 1993; Coelli 1995). Most applications of frontier analysis in Asian aquaculture have used the stochastic approach (Sharma and Leung, 1998; Sharma, 1999; Bimbao, *et. al.* 2000; Dey *et al.,* 2000; Gunaratne and Leung, 2001; Sharma and Leung, 2000 a; b; and Irz and McKenzie 2003). Inefficiency models may be estimated with either a one step or a two-step process. For the two-step procedure the production frontier is first estimated and the technical efficiency of each farm is derived. These are subsequently regressed against a set of variables, which are hypothesised to influence the farm's efficiency. A problem with the two-stage procedure is the inconsistency in the assumptions about the distribution of the inefficiencies. In the first stage, the inefficiencies are assumed to be independently and identically distributed is order to estimate their values. However, in the second stage, the estimated inefficiencies distributions are assumed to be a function of number farm specific factors, and hence are not identically distributed unless all the coefficients of the factors are simultaneously equal to zero (Coelli, Rao and Battese, 1998). FRONTIER uses the ideas of Kumbhakar, Ghosh and McKenzie (1991) and Reifschneider and Stevenson (1991) and estimates all of the parameters in one step to overcome this inconsistency. The inefficiency effects are defined as functions of farm specific factors (as in the two stage approach) but they are then incorporated directly into the maximum likelihood estimation.

The stochastic production frontier for sample carp producers of Kolleru Lake area is specified as follows taking into account the model developed and proposed by Aigner *et al.* (1977) and Meeusen and van der Broeck (1977).

$$LnY_i = \beta_0 + \sum_{k=1}^{8} \beta_k LnX_{ki} + V_i - U_i \text{.............................}(1)$$

where subscript i refer to the i-th farm in the sample; Ln represents the natural logarithm; Y is output variable and Xs are input variables (Table 1 & 2). The details of input and farm specific variables are defined in Table 1. βs are unknown parameters to be estimated; V_i is an independently and identically distributed N (0, σ_v^2) random error; and the U_i is a non-negative random variable associated with technical inefficiency in production, which is assumed to be independently and identically distributed the inefficiency is assumed to here a truncated normal distribution with mean, μ_i and variance, σ_u^2 (|N (μ_i, σ_u^2)|). According to Battese and Coelli (1995), the technical inefficiency distribution parameter, μ_i is defined as:

$$\mu_i = \delta_0 + \sum_{m=1}^{25} \delta_m Z_{mi} \text{.......................................}(2)$$

where Zs are 25 farm-specific variables (table 1 & 2) and δs are unknown parameters to be estimated. Since the dependent variable in (2) is defined in terms of technical inefficiency, a farm-specific variable associated with a negative coefficient will have a positive impact on technical efficiency (Sharma and Leung, 2000a)

Socio economic and demographic variables (table 1) are expected to have some impact on technical efficiency. Similarly farm specific variables (table 1) are also expected to have an impact over technical efficiency affecting carp yield. Farm specific technical efficiency of the i-th sample farm (TE_i) is obtained using the relationship.

Antilog and expected value TE_i = exp ($-U_i$) and E($-U_i$)..................................(3)

The prediction of technical efficiencies is based on the conditional expectation of expression in (3) given the model which specifications (Battese and Coelli, 1988). The parameters for the stochastic production frontier model in (1) and those for the technical inefficiency model in (2) were estimated simultaneously with maximum-likelihood estimation (MLE) using an econometric computer program, FRONTIER 4.1 (Coelli, 1994), that estimates the variance of the likelihood function as $\sigma^2 = \sigma_u^2+\sigma_v^2$ and $\gamma = \sigma_u^2/\sigma^2$

σ^2 and r are not respondent estimates $r=1-\sigma u^2/\sigma^2$

Hypothesis test:

The following hypotheses were tested with a generalized likelihood ratio test to ensure that inefficiency effects were absent from the model.

H_0: $\mu = 0$, the null hypothesis specifies that a simpler half normal distribution is an adequate distribution of data.

H_0: $\gamma = \delta_0 = \delta_1 = \ldots\ldots = \delta_{25} = 0$, technical inefficiency does not exist.

H_0: $\gamma = 0$, technical inefficiency is deterministic. Hence explanatory variables in (2) must be included in along with input and other relevant variables.

H_0: $\delta_1 = \delta_2 = \ldots\ldots = \delta_{25} = 0$, technical inefficiency effects follows a standard truncated-normal distribution with no technical inefficiency effects.

The null hypothesis can be tested using the generalized likelihood -ratio statistic, λ, given by:

$\lambda = -2\,[\mathrm{Ln}\,\{L\,(H_0)\} - \mathrm{Ln}\,\{L\,(H_1)\}]$ ....(4)

where L (H_0) and L (H_1) denote the values of likelihood function under the null (H_0) and alternative (H_1) hypotheses respectively. If the given null hypothesis is true, λ has approximately χ^2 distribution or mixed χ^2 distribution under $\gamma = 0$ (Coelli, 1995).

4. Results

4.1. Maximum livelihood estimates and test of hypotheses

The hypothesis that technical inefficiency effects have a half-normal distribution with mean zero was rejected. Likewise the hypothesis that there is no technical inefficiency was also rejected. The hypothesis that technical inefficiency is deterministic is rejected. The null hypothesis that all the parameters in the technical inefficiency model except the constant term are zero i.e., the technical inefficiency effects have the same truncated normal distribution was also rejected. The likelihood ratio tests indicate that technical inefficiency effects are significant in explaining the variation in production of the carp farms Kolleru Lake. Maximum-likelihood estimates of stochastic production frontier and technical inefficiency models for carp production in Kolleru Lake, Andhra Pradesh, India is presented in Table 3 and generalized Hypothesis tests for model specification and statistical assumptions is presented in Table 4.

The slope coefficient and output elasticities of inputs of the physical factors, excepting water management cost, were as expected. The elasticity of output for feed was highest (0.2001) (P<001) followed by organic manure (0.1411) (P<001). This is indicative of the fact that both feed and organic manure have major influence on yield. Inorganic fertilizer, stocking weight, labour and lime were not statistically significant. Both chemicals and drugs and water management coefficient showed no impact over yield. σ was x (P=0.01). The estimated values of σ_u^2 and σ_v^2 were found to be 0.008408791 and 0.000006594 respectively indicating that technical inefficiency σ_u^2 is greater than the inefficiency caused by stochastic factor σ_v^2. These values indicate that the differences between the observed (actual) and frontier (potential) output are due to inefficiency and not chance alone. A value for of one indicate the difference attributed to the farmers' less than efficient use of technology i.e., technical inefficiency (Coelli, 1995). In the present case the estimated value of γ, which is the ratio of variance of farm-specific TE to total variance of output, is 0.9992 which suggest that 99.92% of error variation in output among the farms is due to technical inefficiency. This suggests the fact that the random component of inefficiency does not have a significant contribution in the analysis. This result is identical to the observations of Sharma *et al.* 1997, Hjalmarsson *et al.* 1996, Coelli and Battese, 1996, Kalirajan, 1981, Ajibefun *et al.* 1996, Ali and Flinn, 1989 suggesting that technical inefficiency is important in explaining the levels and variations in carp yield in Kolleru Lake. The estimated value of λ was found be 35.71020897 which is positive and statistically significant, confirming an effect of TE.

4.2. Estimation of technical efficiency

The results derived from the econometric estimation indicate that technical efficiency (TE) scores for fish farmers of Kolleru Lake ranged from 0.5051 to 0.9899 with a mean of 0.7260 and modal class being 0.6 to 0.7 (Table-5). Mean technical efficiencies of carp culture ponds of Nepal, India, Bangladesh and Pakistan are reported to be 0.676, 0.790, 0.738 and 0.740 respectively for semi-intensive and intensive carp farming and a 0.597, 0.502, 0.475, 0.625 extensive farming are reported to be (Sharma and Leung, 2000a) and between 0.11-1.0 (Jayaraman, 1998). From this study Over 56% of the farmers operate with efficiency score greater than 0.70 and no farmer operate below the technical efficiency of 0.50.

4.3. Potential for technical efficiency improvement

Our results suggest that an average the farmers of Kolleru Lake area are able to obtain about 72.60% mass of potential output from the given set of production inputs. Therefore, in the short run there is a scope of increasing the productivity of carps by around 27.40% by adopting the technology and techniques used by efficient farms. Further, it is observed that if the average farmer in the sample was to achieve the TE level of its most efficient farmer counterpart then the average farmer could realize 26.7 percent cost saving (1 - [72.60/98.99]). Similarly the most technically inefficient farmer may achieve cost saving of 48.9 percent (i.e., 1 - [50.51/98.99]).

4.4 Determinants of Technical Efficiency

Socio-economic, demographic, farm, environmental and non-physical factors are likely to affect the efficiency of operation of any farm (Kumbhakar and Bhattacharya, 1992; Ali and Choudhury, 1990). Altogether Twenty-five nonphysical socio-economic, demographic and farm specific factors table 1 & 2 were included in the model to test the possible impact and association with the technical efficiency of aquaculture farms operating in the Kolleru lake area of the twenty-five factors considered for technical efficiency model, socio economic factors like religion, age, education, primary occupation, pond size, ownership type, source of water, depth of water, sources of advice taken, method of application of drugs, calamities faced, types of labour, types of feed and disease encountered have a positive impact on technical efficiency (table 3). On the other hand caste, number of children, experience, renovation, asset cost, source of seed, loan, harvesting technique, method of application of feed, periodical netting for biomass checking and types of organic manure showed a negative association with technical efficiency (table 3). Of all the farm specific factors showing positive impact mentioned above only pond size and source of water were found significant at ($p<0.05$) and types of labour, religion, method of application of drug, primary occupation and calamities are found significant at ($p<0.10$). Likewise of the variables showing negative impact only periodical netting for biomass checking is found highly significant ($p<0.01$), source of seed and harvesting technique and caste were found significant at lower level of significance ($p<0.10$) demonstrating the influence of these factors in efficiency of carp culture. Indepth analysis of impact of pond size on TE indicates that a consistent increase in mean TE with the size of the farm showing highest TE value of 0.7552 for aquaculture farms of size greater that 10 ha. (table 6)

This is at par with the popular belief that larger farms are capable to capture the economics of size and operates at higher efficiency levels compared to those of marginal and small farms. Pond area is estimated to have positive impact indicating large operations are technically more efficient than smaller ones (Iinuma *et al.*, 1999). Pond size is reported to have positive relationship on productivity (Roy *et al.*, 2002b). Water is the primary component of all aquacultural activities and its importance has been reflected through significant positive impact on TE. Farmers those who have set up their farms by the side of irrigation canals facilitating regular intake and periodical discharge of water were found to operate with significantly higher mean technical efficiency of 0.7287 compared with those using water from other sources with a mean TE of 0.6648 ($F_{cal}=5.57>F_{crit}=3.88$).

Types of labour employed also showed positive impact on TE. Farms employing permanent labour could utilize their services for timely operation of various farm activities reflecting significantly higher mean TE of 0.7354 compared with those who employed temporary labour which a mean TE of 0.6957 ($F_{cal}=10.11>F_{crit}=3.88$). The majority of farms were operated by Hindu's who are influential and economically sound having an advantageous position for securing resources for investment resulting in significantly higher TE of 0.7275 compared to 0.6207 of other religions (F_{cal}=5.33>F_{crit}=3.88). The method of application of drug has also shown a positive impact over TE. To combat disease, drugs were used either diluted in or sprayed on water or mixed with feed shown an estimated TE of 0.7292, 0.7237 and 0.6284 respectively (F_{cal}=1.62<F_{crit}=3.04). Primary occupation was also found to be significant at a 10 percent level indicating that farmers doing fish farming as the main activity showed a significantly higher mean of 0.7359 compared to those of others, showing a mean TE of 0.6896 (F_{cal}=13.02>F_{crit}=3.88). Impact of calamity on TE was found significant (p<0.10). One possible explanation may be that Kolleru lake is located in the flood and cyclone prone areas of Krishna and Godavari districts of Andhra Pradesh. As a result farms that were affected by flood water are automatically flushed out of organic load and metabolic wastes accumulated due to regular use of organic inputs. Calamity showed higher mean TE of 0.7276 compared to a mean TE of 0.7210 of farms not affected by calamities during the last five years of operation, although the difference is not statistically significant (F_{cal}=5.57<F_{crit}=3.88)

In Kolleru lake area, about 86% of farmers resort to periodical netting for biomass checking, which has shown a significant negative impact on TE. This may be due to high turbidity for a prolonged period from frequent meeting. Mean TE of 0.7898 of farms without periodical netting was significantly higher than the TE of 0.7157 by farms with periodical netting (F_{cal}=25.18 > F_{crit}=3.88). Contrary to expectation, source of seed showed a negative impact on TE at 10 percent level of significance thereby indicating that the on-farm seed production system does not improve the efficiency of aquaculture farms. This may be due to seed from outside be of superior quality. From survey it is found that about 66% of sample farmers stocked ponds with seed from outside sources. A mean TE of 0.7351 was observed from farms stocked seed from outside compared with a mean TE of 0.7215 for farms with own seed. Differences in mean of the two groups well not statistically significant (F_{cal}=1.42 < F_{crit} = 3.88). Harvesting technique showed a negative impact on TE. It revealed that 95% of sample farmers resort to single stocking and single harvesting, with a mean TE of 0.7631, a mean of 0.7241 from farm practicing single stocking and multiple harvesting or multiple stocking and multiple harvesting technique was observed. (F_{cal}=2.47<F_{crit} = 3.38). In Kolleru lake, general caste farmers dependent on hired labour thereby showing negative impact on TE possessed area majority of farms. Mean TE for SC, OBC and General farmers were 0.7333, 0.7195 and 0.7276 respectively (F_{cal}=0.33<F_{crit} = 3.04).

5. Conclusion and Policy Implications

Enhancement of productivity is one of the most important aims of carp culture in India that contributes significantly towards food and nutritional security. The present analysis of data has given attention and focused on the issues of productivity enhancement through improvement of TE of the carp farming system with available resources and technology for carp farmers of Kolleru Lake area, the lake is popularly known as Carp Pocket of India for regularly supplying quality protein, particularly to eastern states of India. The stochastic frontier estimate demonstrates an average technical efficiency of 72.60% indicating that there is scope for increasing the output by 27.40% without increasing the levels of inputs. The use of stochastic frontier analysis enables ranking of the carp culture farms based on efficiency of yield of fish, enabling identification of the most efficient and inefficient farms across various carp farming system operating in the Kolleru lake area. The efficient farms may serve as a model farm for improving the efficiency of carp production in the kolleru lake area as well as throughout the whole country where the mean technical efficiency is reported to be lower (Sharma and Leung, 2000b). The empirical results suggest that there are significant possibilities to increase efficiency levels by extending culture operation in large size ponds, provision for supply of fresh water, application of drugs after diluting in water, employing labour on a permanent basis, leasing out the water bodies to farmers whose primary occupation is fish farming, educating the farmers to procure seed from reliable sources, proper harvesting techniques and avoiding periodical netting. From the policy point of view, the above farm specific variables were found to be most promising areas for action, suggesting that policy makers should consider those factors for possible programs at enhancing the productivity without any additional input use. Further studies involving allocation of resources, cost efficiencies are needed.

Acknowledgements

Authors are thankful to Indian Council of Agricultural Research for sponsoring the project and financial assistance for carrying out the investigation. They also grateful to all the Ex-& present Directors of Central Institute of Freshwater Aquaculture for providing necessary facilities and encouragement during the course of study.

Reference

Aigner, D.J., Lovell, C.A.K., Schmidt, P.J., 1977. Formulation and estimation of stochastic frontier production function models. J. Econometrics 6, 21-37.

Ali, M., Chaudhury, M. A., 1990. Inter-regional farm efficiency in Pakistan's Punjab: A frontier production function study. Journal of Agricultural Economics, 41(1), January 62-74.

Anjaneyulu, Y., 2003. Assessment of environmental quality of Kolleru lake and strategic management plans. In: Prasad, M. K. and Y. Anjanelu. B. S. Publication (Eds.), Lake Kolleru environmental status (past and present), Hyderabad-95, 3-15.

Anon, 2002. Final Report of the project 'Technical innovations in aquaculture and its sustainability of farming system in Andhra Pradesh', Central Institute of Freshwater Aquaculture, Kausalyagnaga, Bhubaneswar, 72.

Anon, 2006 a. Pamphlet, National Fisheries Development Board. Department of Animal Husbandry, Dairying and Fisheries, Ministry of Agriculture, Govt. of India, Block 401-402, Maitri Vihar, Huda Commercial Complex, Ameerpet, Hyderabad-500038, Andhra Pradesh, India.

Anon, 2006 b. Inaugural address of Sri Sharad Pawar, Hon'ble Union Minister for Agriculture, Consumer Affairs, Food and Public Distribution. 9th September 2006, National Fisheries Development Board, Hyderabad, India.

Ajibefun, I.A., Battese, G.E., Daramola, A.G., 1996. Investigation of factors influencing the technical efficiencies of smallholder croppers in Nigeria, CEPA working papers, No. 10/96, Department of Econometrics, University of England, Armidale, 19.

Ali, M., Flinn, J.C., 1989. Profit efficiency among basmati rice producers in Pakistan Punjab. American Journal of Agricultural Economics, 71, 303-310

Battese, G.E., Coelli, T.J., 1992. Frontier production function, technical efficiency and panel data: with application to paddy farmers in India. J. Productiv. Anal., 3, 153-169.

Battese, G.E., Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. Journal of Econometrics, 38, 387-399.

Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier function for panel data. Empirical Economics 20, 325-332.

Bimbao, G. B., Paraguas, F. J., Dey, M. M., Eknath, A. E., 2000. Socio-economics and production efficiency of Tilapia Hatchery Operations in Philippines. Aquaculture Economics and Management, 4(1) and (2), 31-46.

Bravo-Ureta, E., Boris, E., Antonio, Pinheiro, E., 1993. Efficiency analysis of developing country agriculture: A review of the frontier function literature. Agricultural and Resource Economics Review 22(1), 88-101.

Chacko, P.I., Abram, J.G., Andal, R., 1952. A Survey of the flora, fauna and fisheries of the Collair lake. Ind. Com. J. Madras, 1, 274-280.

Chaudhuri, H., Chakrabarty, R. D., Sen, P. R., Rao, N. G. S., Jena, S., 1975. A new high in fish production in India with record yields by composite fish culture in freshwater ponds. Aquaculture, 6(1975), 343-355.

Coelli, T.J., 1994. A guide to FRONTIER version 4.1: a computer programme for stochastic frontier production and cost function estimation. Department of Econometrics, Univ. of New England, Australia.

Coelli, T.J., 1995. Estimators and hypothesis tests for stochastic frontier function: a Monte Carlo analysis. Journal of Productivity Analysis 6, 247-268

Coelli, T.J., Rao, D.S.P., Battese, G.E., 1998. An introduction to efficiency and productivity analysis. USA: Kluwer Academic Publishers, USA.

Coelli, T. J., Battese, G.E., 1996. Identification of factors which influence the technical inefficiency of Indian Farmers, Australian Journal of Agricultural Economics, 40(2), 103-128.

Dey, M. M., Paraguas, F. J., Bimbao, G.B., Regaspi, P.B., 2000. Technical efficiency of Tilapia grow out pond operations in the Philippines. Aquaculture Economics and Management, 4(1) and (2), 33-47.

Dutt, S., 1976. On the Fish and Fisheries of Lake Kolleru, Andhra Pradesh. Mem. Soc. Zool., Guntur, 1, 17-27.

FAO, 2005. Aquaculture production, 2003. Yearbook of fishery statistics - Vol.96/2. Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO, 2006. State of World Aquaculture: 2006. FAO Fisheries Technical Paper. FAO. Rome.

Farrell, M.J., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society, series A120, Part 3, 253-281.

Ghosh, A. K., Paul, R. N., Rama Krishnayya, M., Subba Rao, K., Murthy, M. B. R., 1990. Fish health problems in Kolleru Lake, Andhra Pradesh. J. Inland Fish Soc. India, 22(1) and (2), 980-91.

Gopal Krishnayya, Lal, A. K., 1990. Eutrophication, its causes and consequences with particular reference to Kolleru Lake. Recent trends in Limnology: 145-149.

Greene, W.H., 1993. The econometric approach to efficiency analysis, In Fried, H.O., Lovell, C.A.K. and Schmidt, S.S. (Eds), The Measurement of Productive Efficiency, Oxford University Press, New York, 68-119.

Gunaratne, L.H.P., Leung, P. S., 2001. Asian Black Tiger Shrimp Industry: A meta-production frontier analysis, In P.S. Leung, K. R. Sharma (Eds.). Economics and Management of Shrimp and Carp Farming in Asia, Network of Aquaculture Centres in Asia-Pacific, Bangkok, Thailand.

Hjalmarsson, L., Kumbhakar, S.C., Heshmati, A., 1996. DEA, DFA and SFA: A Comparison. Journal of Productivity Analysis, 7,303-327.

Iinuma, M., Sharma, K. R., Leung, P. S., 1999. Technical efficiency of carp pond culture in peninsula Malaysia: an application of stochastic production frontier and technical inefficiency model. *Aquaculture* 175, 199-213.

INCOR, 1983. Ecodevelopment of Kolleru lake (A wetland ecosystem) status, position and approach document, submitted by INCOR, Department of Geo-engineering, Andhra University, Waltair.

Irz, X., McKenzie, V., 2003. Profitability and technical efficiency of aquaculture systems in Pampanga, Philippines, Aquaculture Economics and Management, 7(3) and (4), 195-211.

Jayaraman, R., 1998. Economics and technical efficiency in carp culture in Thanjavur District, Tamil Nadu State, India. In: Eide, A., Vassdal, T. (Eds.), Proceedings of 9th International Conference of the International Institute of Fisheries Economics and Trade, Trade, Tromso, July 8-11, 1998. Norwegian College of Fishery Science, Tromso, Norway, 71-82.

Kalirajan, K., 1981. An econometric analysis of yield variability in paddy production. Canadian Journal of Agricultural Economics, 29, 167-80.

Khasim Imam, D., Rao, C. C. Panduranga, 2003. Status of heavy metal concentrations in fishes from Lake Kolleru. In: Prasad, M. K. and Y. Anjanelu, (Eds.), Lake Kolleru Environmental Status (Past and Present) B. S. Publication, Hyderabad-95, 41-47.

Kirkley, J.E., Squires, D.E., Strand, I.E., 1995. Assessing technical efficiency in commercial fisheries: The Mid-Atlantic Sea Scallop Fishery. Am. J. Agric. Econ., 77(3), 686-97.

Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly equality and inequality restrictions. Econometrica 54, 1243-1248.

Kumbhakar, S.C., Bhattacharya, A., (1992). Price distortion and resource use efficiency in Indian agriculture: a restricted profit function approach. Review of Economics and Statistics, 74, 231-239.

Kumbhakar, S.C., Ghosh, S., Mcguckin, J., 1991. A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. J. Bus. Econ. Statist. 9, 279-286.

Madhavi, 2005. Kolleru: A Choking Lake. Outlook. The weekly Magazine. December 19, 2005, 64 pp.

Meeusen, W., Van der Broeck, J., 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Errors. Int. Econ. Rev., 18: 435-444.

Murthy, V. S. R., 1973. Studies on the taxonomy of fishes of the family cyprinid and on some aspects of biology of Barbs (Puntius sarana) (Hamilton-Buchanan, 1822), from lake Kolleru (Andhra Pradesh), India. Ph. D. Thesis, Andhra University, Waltair.

Padmavathi P., M. K. Durga Prasad, 2003. Lake Kolleru vis-à-vis the fish ponds – a study on the water quality and zooplankton. In: Prasad, M. K. and Y. Anjanelu, (Eds.), Lake Kolleru Environmental Status (Past and Present) B. S. Publication, Hyderabad-95. pp. 16-24.

Rama Murthy, M., Swamy, A. S. R., 2003. Limnological aspects of Lake Kolleru. ibid., 25-34.

Rama Rao, D., 1986-88. Development of fisheries in Kolleru Lake of Andhra Pradesh- a case study. Central Institute of Fisheries Education (*ICAR*), Versova Bombay, in partial fulfillment for the award of post graduate diploma in fishery science: 58 pp.

Ramakrishnan, E., 1980. Integrated development of Kolleru lake area. Report submitted by Administrative Staff College of India, Hyderabad to the Department of Irrigation, Government of Andhra Pradesh.

Ramani, K. V., Anjanelu, Y., 2003. Pollution studies of Kolleru Lake by industrial, agricultural, domestic wastes and corrective management plans. In: Prasad, M. K. and Anjanelu, Y. (Eds.), Lake Kolleru Environmental Status (Past and Present). B. S. Publication, Hyderabad-95. , 52-60.

Rao, M. Balaparameswara, Padmaja, K., 2003. Lake Kolleru- A perspective with special reference to insecticide pollution. Ibid., 48-51.

Rao, P.N., Samba Reddy, A., 2003. Prospecting utility of water lettuce (Pistil steatites L) from Lake Keller. Ibid., 61-66.

Rao, Seshagiri, B.V., 2003. The impact of fisheries on Lake Kolleru, a wetland ecosystem. Ibid., 118-123.

Reddy, P. Munirathnam, Suryanarayan, M., 1983. The fisherman of Kolleru Lake: A study of their ecological adaptations. Proceedings International conference on man and environment, organized by Indian Anthropological Society. Calcutta.

Reifschneider, D., Stevenson, R., 1991. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. International Economic Review, 32, 715-723.

Roy, O.K., Saradhi, K.P., Sahoo, K.N., Saha, G.S., 2002. Statistics of carp culture in Eluru and Kaikalur Mandals of Kolleru Lake in Andhra Pradesh, India. In: Ayyappan, S., Jena, J. K., Mohan Joseph, M., (Eds.), The Fifth Indian Fisheries Forum Proceedings AFSIB, Mangalore and AoA, Bhubaneswar, India, 409-411.

Roy A. K., Sahoo, K. N., Saradhi, K.P., Saha, G.S., 2001a. Aquaculture in Kolleru Lake area for socio-economic development. Environment and Ecology, 19(4), 927-931.

Roy, A.K., Sahoo, K.N., Saradhi, K.P., Saha G.S., 2001 b. A socio economic study on aqua culturists of Eluru and Kaikalur Mandals of Kolleru Lake area of Andhra Pradesh. In: Sinha, M., Kumar, Dhirendra, Katiha, P.K., (Eds.), National conference on eco-friendly management of resources for doubling fish production strategies for 21 century held at Central Inland Capture Fisheries Research Institute (CICFRI), Barrackpore, during 22-23 December, 1999. , 176-179

Roy, A.K., Sahoo, K.N., Saradhi, K.P., Saha, G.S., 2002 b. Farm size and aquaculture productivity relationship. Asian Fisheries Science 15(2000), 129-134, Asian Fisheries Society, Manila, Philippines

Roy, A. K., Saradhi, K. P., Saha, G.S., Sahoo K.N., Antony A., 2004 a. CD-ROM Database on socio-economic aspects of carp culture in Kolleru lake. In: Roy, A.K., A. Antony and M. Rout (Eds.), Bioinformatics and Statistics in Fisheries Research. Central Institute Of Freshwater Aquaculture, Kausalyaganga, Bhubneswar, India, 119-123.

Roy, A. K., Saradhi K. P., Sahoo K. N., Saha G. S., 2004 b. Development of sampling frame of cultivable water area in Kolleru Lake of Andhra Pradesh. In: Roy, A.K., A. Antony and M. Rout (Eds.), Bioinformatics and Statistics in Fisheries Research. Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar, India, 353-356.

Roy, A. K., Saha, G. S., Sarangi, N., 2006. Socio-economic, educational and demographic profile of fish farmers of Kolleru Lake, Andhra Pradesh. In: Compendium of lectures on National Workshop Cum Training on Bioinformatics and Statistics in aquaculture Research, held during 7 - 10 February, 2006, Central Institute Of Freshwater Aquaculture, Kausalyagnaga, Bhubaneswar-751002, India.

Sahoo, K.N., Roy A.K., Saradhi, K.P., Saha, G.S., 2000. Break even analysis of carp culture practice in Eluru Mandal - A Case Study, In: Proc. Fifth Indian Fisheries Forum, held at Central Institute Of Freshwater Aquaculture, Bhubaneswar during 17 -20 Jan 2000, pp-146.

Sahoo, K. N., Roy, A. K., Saradhi, K. P., Saha, G. S., 2001a. A study on marketing channels of fish produce of Kolleru Lake. Journal of Fisheries Economics and Development, 4(1), 35-44.

Saradhi, K.P., Roy A.K., Sahoo K.N., Saha G.S., 2002. On socio economic database management in Aquaculture. In: Ayyappan, S., Jena, J. K., Mohan Joseph, M., (Eds.), The Fifth Asian Fisheries Forum. Proceedings AFSIB, Mangalore and AoA, Bhubaneswar, India, 413-415.

Sarangi, N., Kumaraiah, P., Rangacharyulu, P. V., Giri Bandla, S., 2004. Status of freshwater aquaculture in Krishna - Godavari Delta - A Profile. Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar-751002, pp 29.

Seshavatharam, V., Dutt, B.S.M., 1978. Studies on ecology and weed biology of Kolleru Lake - A technical report of CSIR Research Scheme.

Sharma, K.R., Leung, P., Zaleski, H.M., 1997. Productive efficiency of the swine industry in Hawaii: stochastic frontier vs. data envelopment analysis, Journal of Productivity Analysis, 8,447-459.

Sharma, K. R., Leung, P. S., 1998. Technical efficiency of carp production in Nepal: An application of Stochastic Frontier Production Function Approach, Aquaculture Economics and Management. Vol.2, No. 3, pp. 129-140.

Sharma, K.R., Leung, P., 1999. Technical efficiency of the longline fishery in Hawaii: an application of a stochastic production frontier. Mar. Resource Econ., 13, 259-274.

Sharma, K. R., 1999. Technical efficiency of carp production in Pakistan, Aquaculture Economics and management, Vol.3, No. 2, 131-141.

Sharma, K. R., Leung, P. S., 2000 a. Technical efficiency of carp pond culture in South Asia: An Application of a stochastic meta-production frontier model, Aquaculture Economics and Management. Vol. 4, Nos. 3 and 4, pp. 169-189.

Sharma, K.R., Leung, P. S., 2000 b. Technical efficiency of carp production in India: A stochastic frontier production function analysis, Aquaculture Research, (31), 937-958.

Vidyanath, V., 2003. Planning atlas for integrated resource development and ecozoning of lake Kolleru. In: Prasad, M. K. and Y. Anjanelu. B. S. Publication (Eds.), Lake Kolleru Environmental Status (Past and Present) Hyderabad-95, 35-38.

Table 1. Definitions of variables and measurement units for empirical model

Variables	**Description**
Output (Y) = Yield	Aggregated quantity of fish yield (in ton/ha/year)
Input	
Organic Manure	Amount of organic manure (cowdung, poultry dropping) used in carp production (in ton/ha)
Inorganic fertilizer	Amount of inorganic fertilizers used in carp production (in kg/ha)
Lime	Quantity of Lime used during culture duration (kg/ha)
Stocking weight	Number of fishes stocked multiplied by average weight (kg/ha)
Chemicals and Drugs	Amount spent during culture operation (Rs./ha)
Feed	Total dry weight of feed ingredients consisting of deoiled rice bran and mustard oil cake/groundnut oil cake/cotton seed cake and soyabean applied in carp ponds (in ton/ha)
Labour	Hired labour engaged in carp culture for stocking, harvesting, feeding, watch and word, etc. (mandays/ha)
Water management	Water management cost included expenditure incurred during initial intake of water and periodical replenishment for loss due to seepage and evaporation (Rs./ha) to maintain level of water suitable for aquaculture.
Technical efficiency (TE)	
Religion	Value 1 if Hindu, 0 if otherwise
Caste	Value 1 if SC, 2 if OBC and 3 if general caste
Age	Age of the farmers (in years)
No. of children in family	Children of the farmer in number
Education	Number of years in school
Experience in carp culture	Number of years the farmer is engaged in fish farming
Primary occupation	Value 1 if primary occupation in fish farming, 0 otherwise
Pond size	Area of pond in hectares
Pond renovation	Value 1 if renovation done or 0 if otherwise
Asset cost	Value (Rs./ha) of various types of assets like water pump, boat, net, godown, etc.
Ownership type	Value 1 if owner operated, 0 if otherwise

Source of water	Value 1 if taken from adjoining irrigation canal, 0 if otherwise
Depth of water	Average depth (m) of water of the culture pond
Source of seed	Value 1 if reared in own farm, 0 if otherwise
Sources of advice taken	Value 1 if govt. sector, 2 if fish doctor, 3 if self and 4 if fellow farmer
Loan	Value 1 if taken from bank, 0 if otherwise
Harvesting technique	Value 1 if adopted single stocking and single harvesting practice, 0 if otherwise
Method of application of drug	Value 1 if mixed with feed and applied, 2 if sprayed on water, 3 if diluted in water.
Method of application of feed	Value 1 if given in perforated bags hanged in poles and submerged inside pond, 0 if otherwise.
Calamities faced	Value 1 if calamities faced during last five years, 0 if not faced
Periodical netting for biomass checking	Value 1 if netting done for biomass checking, 0 if otherwise
Types of labour	Value 1 if permanent labor engaged, 0 if otherwise.
Types of organic manure	Value 1 if poultry droppings only, 2 if cowdung and poultry dropping
Types of feed	Value 1 if applied cereal grain byproduct (DORB); 2 if applied [DORB and oil seed cake (GNOC, MOC, Sun flower seed cake)]; 3 if applied [DORB, oil seed cake and Protein (Soyabean and Nutra)]; 4 if applied DORB and Soyabean
Disease encountered	Value 1 if fungal only, 2 if bacterial only, 3 if viral only, 4 if parasite only, 5 if fungal and bacterial, 6 if fungal and viral, 7 if viral and bacterial, 8 if fungal, bacterial and viral, 9 if parasite and bacterial, 10 if parasite and viral and 11 if parasite, fungal and bacterial

DORB: De oiled Rice Bran; CSC: Cotton Seed Cake; GNOC: Ground Nut Oil Cake; MOC: Mustard Oil Cake

Table 2. Summary statistics of variables involved in the stochastic production frontier and technical inefficiency models for carp pond culture in Kolleru Lake, Andhra Pradesh, India.

Variable	Mean	SD	Minimum	Maximum
Output (ton/ha)	10.630	1.470	7.400	14.800
Organic manure (ton/ha.)	22.790	5.100	3.700	37.100
Inorganic fertilizer (kg/ha.)	1507.250	501.640	617.750	3953.600
Lime (kg/ ha.)	1075.130	379.700	494.200	2471.000
Stocking weight (kg/ ha.)	1939.000	512.450	474.430	3059.100
Chemicals and Drugs (Rs./ ha.)	3161.61	1409.95	432.00	8159.00
Feed (ton/ha.)	27.650	5.760	13.700	44.500
Labour (mandays/ha.)	88.400	53.300	5.0	242.0
Water management cost (Rs./ ha.)	9360.00	6127.00	634.00	21004.00
Farm-specific				
Religion (0 or 1)	.99	.116	0	1
Caste (1 or 2 or 3)	2.50	.698	1	3
Age (year)	42.07	5.703	30	80
No of children (no.)	2.36	1.016	0	5
Education (no. of years in school)	10.18	1.952	9	17
Experience (years)	10.42	5.141	1	20
Primary occupation (0 or 1)	0.79	0.41	0	1
Pond size (ha.)	8.7	8.6	0.8	61.0
Renovation (0 or 1)	.83	.37	0	1
Asset cost (Rs./ha.)	12873.12	9374.50	1060.00	67950.00
Ownership type (0 or 1)	0.92	0.27	0	1
Source of water (0 or 1)	0.96	0.19	0	1
Depth of water (meter)	2.44	0.35	1.68	3.35
Source of seed (0 or 1)	0.66	0.475	0	1
Advice taken from (1 or 2 or 3 or 4)	3.38	0.85	1	4
Loan (0 or 1)	0.30	0.46	0	1
Harvesting technique (0 or 1)	0.95	0.22	0	1
Method of application of drug (1 or 2 or 3)	2.57	.51	1	3

Method of application of feed (0 or 1)	0.97	0.17	0	1
Calamities (0 or 1)	0.76	0.425	0	1
Periodical netting (0 or 1)	0.86	0.348	0	1
Types of labor (0 or 1)	0.76	0.425	0	1
Types of organic fertilizer (1 or 2)	1.48	0.501	1	2
Types of feed (1 or 2 or 3 or 4)	2.33	0.58	1	4
Disease encountered (1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11)	3.44	2.089	1	11

Table 3. Maximum-likelihood estimates of stochastic production frontier and technical inefficiency models for carp production in Kolleru Lake, Andhra Pradesh, India.

	Parameter	Coefficient	Standard-error	T-ratio
Stochastic production frontier model				
Constant	β_0	1.2374793***	0.19531434	6.3358343
Organic manure (X_1)	β_1	0.14112453***	0.037727305	3.7406470
Inorganic fertilizer (X_2)	β_2	0.024032115	0.023807879	1.0094186
Lime (X_3)	β_3	0.033957713	0.026476323	1.2825691
Stocking weight (X_4)	β_4	0.026828252*	0.019799683	1.3549839
Chemicals and Drugs (X_5)	β_5	-0.014569666	0.017009698	-0.85655054
Feed (X_6)	β_6	0.20010391***	0.036975060	5.4118616
Labor (X_7)	β_7	0.011732542	0.011203558	1.0472157
Water management (X8)	β_8	-0.022573018*	0.011707627	-1.9280609
Technical inefficiency model				
Constant	δ_0	0.55002287	0.60348210	0.911415
Religion (Z_1)	δ_1	-0.085277684*	0.050847024	-1.6771421
Caste (Z_2)	δ_2	0.013630104*	0.0097874157	1.3926152
Age (Z_3)	δ_3	-0.00051804104	0.0013822222	-0.3747885
No of children (Z_4)	δ_4	0.0041073793	0.0084882335	0.4838909
Education (Z_5)	δ_5	-0.0016462576	0.0040847498	-0.4030253
Experience (Z_6)	δ_6	0.00022949786	0.0022022899	0.10420874
Primary occupation (Z_7)	δ_7	-0.027911207*	0.022389016	-1.2466473
Pond size (Z_8)	δ_8	-0.0026728321**	0.0011043401	-2.4202980
Renovation (Z_9)	δ_9	0.023636240	0.019656041	1.2024924

Asset cost (Z_{10})	δ_{10}	0.00000041221	0.0000010251	0.4020827
Ownership type (Z_{11})	δ_{11}	-0.037148235	0.033218818	-1.1182889
Source of water (Z_{12})	δ_{12}	-0.097534707**	0.040034542	-2.4362639
Depth of water (Z_{13})	δ_{13}	-0.0008716901	0.023454639	-0.0371649
Source of seed (Z_{14})	δ_{14}	0.030013137*	0.016411887	1.8287439
Sources of Advice taken (Z_{15})	δ_{15}	-0.0045832209	0.010281450	-0.4457757
Loan (Z_{16})	δ_{16}	0.021655156	0.024542989	0.8823357
Harvesting technique (Z_{17})	δ_{17}	0.059316720*	0.040854522	1.4519010
Method of application of drug (Z_{18})	δ_{18}	-0.015432842*	0.010162719	-1.5185742
Method of application of feed (Z_{19})	δ_{19}	0.0033167997	0.043048858	0.0770473
Calamities faced (Z_{20})	δ_{20}	-0.029885695*	0.018107579	-1.6504523
Periodical netting for biomass checking (Z_{21})	δ_{21}	0.068515871**	0.025246614	2.7138638
Types of labour (Z_{22})	δ_{22}	-0.042361853*	0.030499051	-1.3889565
Type of organic manure (Z_{23})	δ_{23}	0.0091244716	0.013217207	0.6903479
Type of feed (Z24)	δ_{24}	-0.019243984	0.016299137	-1.1806750
Disease encountered (Z_{25})	δ_{25}	-0.00026916853	0.0040550140	-0.0663799
	σ^2	0.0084153852	0.00077490850	10.8598440
	γ	0.99921654	1.3397727	0.74581048
	Log likelihood function	214.76039		
	LR test	63.053464		
	Mean TE	0.72608099		

***, **, * Significant at 1, 5 and 10 percent level of significance respectively

Table 4: Generalised Hypothesis tests for model specification and statistical assumptions

Null Hypothesis (H_0)	**Log (likelihood)**	**Test statistic (λ)**	**Critical value ($\chi^2_{0.05}$)**	**Decision**
$\mu = 0$	210.79915	7.92248	3.84	Reject H_0
$\gamma = \delta_0 = \delta_1 = \ldots\ldots = \delta_{25} = 0$	158.77339	111.974	39.53[a]	Reject H_0
$\gamma = 0$	198.16121	33.19836	3.84	Reject H_0
$\delta_1 = \delta_2 \ldots\ldots = \delta_{25} = 0$	183.23366	63.05346	37.63	Reject H_0

a : $\chi^2_{0.05}$ is obtained from Kodde and Plam (1986, P.1246)

Table 5. Frequency distribution of technical efficiency estimates for a stochastic frontier model.

Efficiency score class	**Number of farmers**	**% of farmer**	**Cumulative frequency of farmer**
0.50 - 0.60	6	2.71	17
0.60 - 0.70	90	40.72	125
0.70 - 0.80	88	39.82	199
0.80 - 0.90	30	13.58	218
0.90 - 1.0	7	3.17	221
Total	221	100.00	--

Table 6: Relationship between pond size on TE level of carp farm in Kolleru Lake, Andhra Pradesh, India

	Marginal (< 1 ha)	**Small (1-2 ha)**	**Semi-medium (2-4 ha)**	**Medium (4-10 ha)**	**Large (>10 ha)**
Mean TE	0.63525835	0.676256767	0.71594478	0.721440235	0.755257403
Standard Deviation	0.017217	0.077395	0.067207	0.07777	0.083428
Coefficient of Variation (%)	2.710266	11.44464	9.387159	10.77982	11.04634

MACHINE LEANING – AN EVOLVING RESEARCH TREND IN BIOINFORMATICS

Sushma Rani Martha[1]

The discipline of bioinformatics, which uses computer technology to provide answers to biological questions, has been expanding in scope and utility for the past decade. Increasing numbers of research groups have been investing in bioinformatics infrastructure to aid in the research process. The cross-disciplinary nature of bioinformatics entails co-evolution with other biomedical disciplines, whereby some bioinformatics applications become popular in certain disciplines and, in turn, these disciplines influence the focus of future bioinformatics development efforts. We observe here that the growth of computational approaches within various biomedical disciplines is not merely a reflection of a general extended usage of computers and the Internet, but due to the production of useful bioinformatics databases and methods for the rest of the biomedical scientific community. Use of computational methods has become increasingly prevalent across biomedical disciplines over the past three decades, while use of databases and the Internet have been rapidly increasing over the past decade. Research and development in two important areas are impacting the science and technology. High throughput sequencing and molecular imaging technologies marked the beginning of a new era for modern translational medicine and personalized healthcare. The impact of having the human sequence and personalized digital images in hand has also created tremendous demands of developing powerful supercomputing, statistical learning and artificial intelligence approaches to handle the massive bioinformatics and personalized healthcare data, which will obviously have a profound effect on how biomedical research will be conducted toward the improvement of human health and prolonging of human life in the future.

Bioinformatics aims to promote interdisciplinary and multidisciplinary education and After availability of many genomic sequences the researcher's started focusing on important research fields including gene expression data analysis and applications, high-throughput genome mapping, sequence analysis, gene regulation, protein structure prediction, disease prediction by machine learning techniques, systems biology, database and biological software development.

Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge. This capacity to learn from experience, analytical observation, and other means, results in a system that can continuously self-improve and thereby offer increased efficiency and effectiveness.

It deals with the changes in systems that perform tasks associated with *artificial intelligence (AI)*. Such tasks involve recognition, diagnosis, planning, robot control, prediction, etcAs robot technology advances, we are approaching the day when robots will be deployed prevalently in uncontrolled, unpredictable environments: the proverbial 'real world.' As this happens, it will be essential for

[1] Correspondence: *Sushma Rani Martha;* Assistant Professor, Department of Bioinformatics, CPGS, OUAT, Bhubaneswar

these robots to be able to adapt autonomously to their changing environment. For a robot to learn to improve its performance based entirely on real-world environmental feedback, the robot's behavior specification and learning algorithm must be constructed so as to enable data-efficient learning.

Machine Learning is a scientific field addressing the question 'How can we program systems to automatically learn and to improve with experience?' We study learning from many kinds of experience, such as learning to predict which medical patients will respond to which treatments, by analyzing experience captured in databases of online medical records. It also includes the study of mobile robots that learn how to successfully navigate based on experience they gather from sensors as they roam their environment, and computer aids for scientific discovery that combine initial scientific hypotheses with new experimental data to automatically produce refined scientific hypotheses that better fit observed data. To tackle these problems we develop algorithms that discover general conjectures and knowledge from specific data and experience, based on sound statistical and computational principles. Machine learning approaches to natural language processing problems such as information retrieval, document classification, and information extraction have developed rapidly over recent years. Even more recently, the joint analysis of text and images has become a significant focus for machine learning.

INFORMATION MANAGEMENT FOR FISHERIES POLICY DEVELOPMENT - ISSUES AND OPTIONS

Pradeep K. Katiha[1]

1 Introduction

Last few decades witnessed rapid changes in world fisheries and India is in no way exception (Ahmed *et al.*, 1999). Drastic changes in the management of fisheries and new technologies have resulted in enhanced access and significant expansion of effort and production in fisheries sector. World fish production grew to a record level of 141 million t in 2004 (FAO, 2007) from only 20 million t in the early 1950s. For India it reached from 0.75 million t to 6.8 million t in 2007-08. Against the backdrop of rapid growth and expansion of global fisheries production and trade, the structural characteristics of developing countries with respect to property rights, resource access and barriers to market entry have resulted in many instances of environmental degradation and produced significant inequality in the distribution of benefits between countries and between various groups within countries. Over the past two to three decades, fisheries issues have emerged from being an obscure sectoral concern or primarily a welfare consideration for few coastal and inland people to an important growth sector having a significant role in economic development and food security in many developing countries (Ahmed *et al.*, 1999a). While this is well known to many public and private actors of the developing world, it is still largely ignored by policymakers in those countries. So far it has been largely limited to scientific discussion among fisheries biologists. This may explain in part why research on fisheries policies and management of information on fisheries in developing countries is still a relatively new field. Only recently has the importance of the actual and potential contribution of fisheries and aquaculture become part of the debate about food security and environmental sustainability in developing countries.

Within the area of interface between fisheries, food security, and environmental sustainability, the fundamentals of achieving economic growth, reducing poverty and protecting natural resources and the environment are the major objectives of fisheries policy. This chapter is primarily organized with a view to soliciting inputs from literature on fisheries policy from developing countries and management of information for development of policies for the sector.

To start with, it is necessary to define policy and how it helps in development process.

2. How can policy help?

A policy is a course of action (explicitly) adopted or pursued to achieve one or a set of objectives, taking into account available opportunities and constraints.

[1] Correspondence: P. K. Katiha, Principal Scientist, Central Inland Fisheries Research Institute, Barrackpore, West Bengal -700120 pkatiha@yahoo.com

Policymaking requires a choice among a set of alternative options on the basis of an objective assessment of their pros and cons. Policies are made either in a proactive mode to create new options and pathways, or in response to changed conditions or public pressure. Accordingly the information on parameters for policy development can provide insights into the issues and prevailing factors affecting them and suggest new options and implementation strategies and their likely consequences (Williams, 1996).

3. Fishery policy typologies and management of information

To have the information required for a fisheries policy, the prerequisite is the type of policy for which information is to be gathered. It could be approached in various ways and a number of typologies may be suggested (Garcia, 1999). For instance, by jurisdictional areas one can group policies in terms of national and international policies, where national policies cover both sectoral and inter-sectoral issues, while international policies deal with trans-boundary or trans-border issues. On the other hand, policies can be specific to a particular production area or system, such as marine fisheries, inland fisheries, and aquaculture. The analysis of the priorities in the three main production systems (marine fisheries, inland fisheries and aquaculture) showed that, when properly expressed, most of the broad priority issues are common to the three systems and only a few specific differences exist. Nevertheless, management of information and research aimed at strengthening and improving policies for fisheries in developing countries need to integrate three main levels of consideration, such as: i) filling the information gap from a disciplinary point of view; ii) providing an understanding of the policymaking process; and iii) identification of the main policy issues and challenges for information management.

In case of fisheries, there are three priority areas for policy development: sustainable development and responsible fishing, food security and capacity building. Accordingly the information should be gathered and managed on the issues and options for each priority area of policy development. These are summarised in following paragraphs.

4. Sustainable development and responsible fishing

Policies that relates mainly to sustainable development and responsible fishing will require information on combination of various issues effecting aquatic systems (Garcia, 1999). These issues are:

- i) conservation of resources and the environment for future generations;
- ii) optimizing economic value (and efficiency) of the sector;
- iii) promoting improved and fair trade;
- iv) regulating access to resources;
- v) optimizing interactions between small-scale and industrial sectors;
- vi) optimizing interactions with other sectors;
- vii) improving governance; and
- viii) accompanying socioeconomic transformation

These may be applicable for different fisheries sectors (Table 1) with different priorities and importance. For sustainable development and responsible fisheries major priority issues are for marine and inland fisheries.

Table 1 Management of information issues related to policies for sustainable development and responsible fishing for different fisheries sectors

Issues	Input information to be gathered and managed on
1. Conservation of the resources and the environment for future generations	• Measures to develop habitat protection and rehabilitation strategies (M, I, A) • Aspects related to develop resource rebuilding strategies for depleted stocks (M, I) • Steps to improve protection of biodiversity (including endangered species, introductions) (M, I, A) • Issues relevant to design precautionary approaches (A) • Conduct of experiment with ecosystem having multi-species management (I,A) • Strategies to develop conservation strategies for water resources (I,A) • Options to develop ecosystem-based management approaches (M,I,A) • Reduction flak in fluctuating resources management (M,I,A) • Factors responsible for pressure on wild seeds (M, I) • Potential for a market driven (demand driven) management system (M, I, A) • Parameters for comparative analysis of management strategies (M, I, A) • Assessment of precautionary development/ management strategies (M, I) o impact/procedures for Prior Consent Procedures (PCPs) o impact/prócedures for Prior Information Procedures (PIPs) ▪ Study impacts of Impact-.based activities and sources of pollution (M, I, A)

Issues	Input information to be gathered and managed on
2.Optimizing economic value and efficiency of the sector	• Issues to develop capacity control and reduction strategies (M, I, A) • Effects of economic subsidies, Incentives and Disincentives (M, I, A) • Development of fisheries rehabilitation programs (M, I, A) • Impact of economic globalization (M, I, A) • Improve natural resources valuation (M, I) • Cost recovery strategies for management (M,I,A) • Factors effecting fish prices (M, I, A)
3. Promoting Improved and fair trade	• Regional and international trade opportunities (M, I, A) • Impact of environmental regulations on trade (non-tariff barriers) (M, I, A) • Measures to increase export value (M, I, A) • Development of quality assurance (M, I, A) • Impacts of trade liberalization (M, I, A) • Comparison of self-sufficiency policies vs. import-export balancing policies (M, I, A) • Parameters to analyse demand elasticities and substitutability (M, I, A)
4. Regulating access to resources	• Competition between artisanal and industrial sectors (M, I, A) • Interactions between national and foreign operators (M, I, A) • Impact of access rules on gender issues (M, I, A) • Rent appropriation (M, I, A) • Impact of access rules on intra-generational equity (M, I, A) • Bases for a rights-based fishery management system (M, I, A) • Rights of access to water resources (M, I, A)

Issues	Input information to be gathered and managed on
5.Optimizing interactions between small-scale and industrial sectors	• Market interactions (e.g., impact on price formation) (M, I, A) • Potential synergies and conflict reduction measures (M, I, A) • Sector dynamics and relationships (labor and capital) (M, I, A) • Differential investment strategies (M, I, A) • Role and impact of fishing agreement with foreign fleets (M, I, A) • Compliance of fishing agreement (M, I, A) • Development of conflict resolution mechanism (M, I, A)
6.Optimizing interactions with other sectors	• Competition between the fishery and other sectors (M, I, A) • Integration of aquaculture with rural development and agriculture (A) • Impacts of national economic and development policies (M, I, A)
7.Improving governance	• Partnership management opportunities/options (M, I, A) • Decentralization programs (M, I, A) • Effectiveness of regional commissions (M, I, A) • Requirements for sectoral integration (M, I, A) • Consequences/modalities for privatization (M,I,A)
8.Accompanying socio-economic transformation	• Dynamics of labor force (M, I, A) • Impact off change on gender issues (M, I, A) • Sectoral demographic studies (growth, migrations, urbanization) (M, I, A) • Promotion of professional organizations (M, I, A) • Forecast impacts of structural adjustments (M,I,A) • Improvement of mechanisms for technology transfer (M, I, A) • Comparison of impacts of export vs. import-oriented policies (M, I, A) • Support to fishers' education (M, I, A) • Appropriate transitional pathways (minimizing social stress) (M, I, A)

Modified from Garcia, 1999, M = marine; I = inland; A = aquaculture

5. Food security

Information collection and management on issues related to food security should focus on:

i) Improving access to food;
ii) Improving use of underused resources;
iii) Reducing waste;
iv) Enhancing natural productivity; and
v) Developing aquaculture

Similar to policies for sustainable development and responsible fisheries, the information to be gathered and managed for policies related to food security may also be applicable for different aquatic ecosystems (Table 2) with different priorities and importance. For food security major priority issues are for all the ecosystems.

Table 2 Management of information for issues related to policies for food security

Research Issues	Input information to be gathered and managed on
1. Improving access to food	• Constraints to access to food (M, I, A) • Consumer preferences (M, I, A) • Development of products (M, I, A) • Improvement of national trade (M, I, A) • Role of international trade in national process, access to food (M, I, A) • Impact of export policies on food availability (and buying power) (M, I, A)
2. Improving use of underused resources	• Underused resources (M, I, A) • Auto-centered development strategies (M, I, A) • Pros and cons of agreement with foreigners (M, I, A) • Fishmeal for feeds vs. food for humans (M, I, A) • Forecast/management of unstable/variable resources (M, I, A)
3. Reducing waste	• Improve gear selectivity (reduce by catch) (M) • Improve use of by catch, reduce discards (better processing) (M) • Reduce other postharvest losses (M, I, A)
4. Enhancing natural productivity	• Impacts on biodiversity and environments (M, I, A) • Impacts of ranching (M, I, A)
5. Developing aquaculture	• Species introduction vs. selective breeding (A) • Land-based vs. water-based aquaculture (A)

	• Relations between water/habitat management and diseases (A) • Resource-use conflicts (A) • Integration with other sectors (A) • integration with agriculture rural development and other sectors (A) • Subsistence vs. commercial aquaculture technologies (A) • Long-term societal impact of developing an enabling environment for aquaculture development (A) • Promoting applied research capacity (A) • Management plans for aquaculture development (A) • Common property resources for aquaculture (A) • Partnerships in managing production systems (A) • Role of governments in aquaculture infrastructures (A) • Public-private interface, seed production, feed supply, etc. (A) • Policy formation and implementation process (A) • Role of NGOs in aquaculture development (A) • Elements of risk in aquaculture

Modified from Garcia, 1999, M = marine; I = inland; A = aquaculture

6. Capacity building

A review of the current level of information and the strength of national institutions concerned with fisheries suggests that significant effort will need to be made to develop policy research and capacity. In particular, developing countries will need substantial support in the following areas:

i) Collection of baseline information;
ii) Improving long term monitoring and forecasts;
iii) Research methodologies and analytical tools; and
iv) Improving policy analysis

The information to be gathered and managed for policies related to capacity building may be applicable for all the aquatic ecosystems. The issues for the same are summarised in Table 3.

Table 3 Management of information for issues related to capacity building in fisheries

Issues	Input information to be gathered and managed on
1.Improving collection of baseline	• Optimization of statistical system • Collection of strategic data (particularly socioeconomics)

information	• Improvement in household level data • Development of sub-sectoral sustainability indicators
2.Improving long-term monitoring and forecasts	• Projections for fish supplies • Assessment and monitoring of wild and enhanced resources • Assessment and monitoring of aquaculture resources (water, land, feeds) • Projection of demand and market trends for fish and fishery products • Computation and analysis of demand elasticity and product substitutability • Analysis of cultural constraints • Interaction of aquaculture and capture fisheries on prices • Analysis of national price trends and price formation • Projection of fisheries management and planning institutional capacity • Projection of environmental protection requirements
3. Development of research tools	• Database and integrated information systems • Models and software (food consumption, availability, etc.)
4.Improving policy analysis	• Development of approaches for performance analysis • Analysis of mechanisms for policymaking (participation, etc.) • Improvement of feedback mechanisms

Modified from Garcia, 1999

From an inter-disciplinary angle, information on fisheries policy should consider among others the following issues (Garcia, 1999):

i) Eco-biological issues: overfishing, rehabilitation, endangered species, discards;

ii) Economic issues: economic incentives and disincentives, trade, resource allocation;

iii) Technological issues: technologies, waste reduction;

iv) Social, cultural, political issues: labour, gender, demography, conflict, access and allocation; and

v) Institutional and legal issues: rights, organizations, legal frameworks, enforcement, trans-boundary negotiations. In relation to the policymaking process, both research in support of policy (methodologies, assessments and option analysis) and research on policy (policy formation, processes and performance) are needed.

Some major policy issues and challenges related to key policy objectives are:

- (i) Food security: supply, access, safety;
- (ii) Sustainability: development and threats to productive systems;
- (iii) Poverty and growth;
- (iv) Human resources development;
- (v) Demography and urbanization;
- (vi) Capacity building: infrastructures, markets;
- (vii) Macro-economic constraints: exchange rates, trade regulations; and
- (viii) Financial resources mobilization and international assistance.

7. Conclusion

Development of any sector largely depends upon the policies followed and their suitability, applicability, implementation, etc. These policies in turn are a function of background information, which forms the basis for their formulation. The insufficient information or lacuna in its collection would lead to low applicability and inefficient implementation of the policy. It will directly influence the development of the sector. Fisheries are in no way an exception. Further, considering the natural renewable multiple use resource, poor socio-economic status of the fisher community, low equity and welfare concerns, poor infrastructure facilities in the sector and low technical know how of fisher community made the policies increasingly important for the sector. For policy development, information may be related to three basic objectives of: sustainable development and responsible fishing, food security and capacity building. Keeping above-mentioned facts in mind, collection of basic information, its analysis and management require greater emphasis to support better policy development for the sector.

Selected Readings

Ahmed, M., C. Delgado and S. Sverdrup-Jensen. 1997. A brief for fisheries policy research in developing countries. ICLARM, Manila, Philippines. 16 p.

Ahmed, M., C. Delgado and S. Sverdrup-Jensen. 1999. The growing need for fisheries policy Research in developing countries, p. 1-4.*In* M. Ahmed, C. Delgado, S. Sverdrup-Jensen and R.A.V. Santos (eds.) Fisheries policy research in developing countries: issues, priorities and needs. ICLARM Conf. Proc. 60, 112 p.

Ahmed, M., C. Delgado, S. Sverdrup-Jensen and R.A.V. Santos (eds). 1999a. Fisheries policy research in developing countries: issues, priorities and needs. ICLARM Conf. Proc. 60, 112 p.

Ahmed, M., C. Delgado and **S.** Sverdrup-Jensen. 1999b. Priority policy research agenda and their implementation, p. 101-103. *In* M. Ahmed, C. Delgado, S. Sverdrup-Jensen and R.A.V. Santos (eds.) Fisheries policy research in developing countries: issues, priorities and needs. ICLARM Conf. Proc. 60, 112 p.

Charles, A. 1998. Fisheries in transition, p. 15-37. Ocean Yearbook 13. The University of Chicago.

FAO (Food and Agriculture Organization of the United Nations). 2007. The state of world fisheries and aquaculture, 2006. FAO, Rome, Italy.

Garcia S. 1999. Setting regional and global priorities targeted to developing countries: discussion summary, p. 95-100. *In* M. Ahmed, C. Delgado, S. Sverdrup-Jensen and R.A.V. Santos (eds.) Fisheries policy research in developing countries: issues, priorities and needs. ICLARM Conf. Proc. 60, 112 p.

Sverdrup-Jensen, S. 1999. Policy issues deriving from the impact of fisheries on food security and the environment in developing countries, p. 73-91. *In* M. Ahmed, C. Delgado, S. Sverdrup-Jensen and R.A.V. Santos (eds.) Fisheries policy research in developing countries: issues, priorities and needs. ICLARM Conf. Proc. 60, 112 p.

Williams, M.J. 1996. The transition in the contribution of living aquatic resources to food security. International Food Policy Research Institute: Food Agric. Environ. Discuss. Pap. 13, 41 p. (An edited version of this paper also appeared in the special 10th Anniversary publication of the Asian Fisheries Society.)

GIS AND ITS APPLICATIONS IN AGRICULTURE - RECENT DEVELOPMENTS

Prachi Misra Sahoo[1]

1.0 Introduction

Geographic Information System (GIS) is a computer based information system used to digitally represent and analyse the geographic features present on the Earth' surface and the events that taking place on it. Geographical information system (GIS) has become of increased significance for environmental planning and assessment mainly because of the need to compare a great number of spatially related data, and because it can be used to couple these spatial data with their attributes and overlay them. It not only provides a visual inventory of the physical, biological, and economical characteristics of the environment, it also allows rational management without complex and time-consuming manipulations. GIS does not provide a definitive answer to a given problem; rather, it generates outputs to a range of input data. It does support the decisions of managers built up with the outputs from the GIS, and perhaps other related material. This paper gives an overview of basics of GIS and its use in agriculture, fishery and aquaculture research, recent developments and future trend.

2.0 Defining GIS

A GIS is an information system designed to work with data referenced by spatial/ geographical coordinates. In other words, GIS is both a database system with specific capabilities for spatially referenced data as well as a set of operations for working with the data. It may also be considered as a higher order map. A Geographic Information System is a computer based system which is used to digitally reproduce and analyse the feature present on earth surface and the events that take place on it. In the light of the fact that almost 70% of the data has geographical reference as it's denominator, it becomes imperative to underline the importance of a system which can represent the given data geographically.

Three perspectives on GIS

a. GIS as a toolbox -> if so then what kind of tools?

- Classification based on functional tasks of GIS:
- Tools for automating *spatial data* (data capture via digitizing, scanning, remote sensing, satellite geo-position system)
- For storing spatial data (data bases and data structures)
- For spatial data management/retrieval
- For analysis (overlay, buffering, proximity, network functions, spatial statistics)
- For display of spatial data and analysis results

[1] **Correspondence**: Dr. (Mrs) Prachi Mishara Sahoo, Scientist (Senior Scale); Division of Sample Survey; Indian Agricultural Statistics Research Institute; New Delhi - 110 012

b. GIS as an Information System

- definition of GIS as a specialized information system stresses "spatially distributed features (points, lines, areas), activities (physical and human-invoked), and events (time)
- Definition of GIS: A system of hardware, software, data, people, organizations, and institutional arrangements for collecting, storing, analyzing and disseminating information about areas of the earth.
- GIS as an approach to Geographic Information Science
- research on GIS (algorithms, analytical methods, visualization tools, user interfaces, human-computer-human interaction)
- research with GIS: GIS as a tool used by many substantive disciplines in their own ways (anthropology, archeology, forestry, geology, engineering, business and management sciences)

3. Components of GIS

GIS constitutes of five key components:
- Hardware
- Software
- Data
- People
- Method

3.1 Hardware

It consists of the computer system on which the GIS software will run. The choice of hardware system range from 300MHz Personal Computers to Super Computers having capability in Tera FLOPS. The computer forms the backbone of the GIS hardware, which gets it's input through the Scanner or a digitizer board. Scanner converts a picture into a digital image for further processing. The output of scanner can be stored in many formats e.g. TIFF, BMP, JPG etc. A digitizer board is flat board used for vectorisation of a given map objects. Printers and plotters are the most common output devices for a GIS hardware setup.

3.2 Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. GIS softwares in use are MapInfo, ARC/Info, AutoCAD Map, etc. The software available can be said to be application specific. When the low cost GIS work is to be carried out desktop MapInfo is the suitable option. It is easy to use and supports many GIS feature. If the user intends to carry out extensive analysis on GIS, ARC/Info is the preferred option. For the people using AutoCAD and willing to step into GIS, AutoCAD Map is a good option. The software for a geographical information system may be split into five functional groups as mentioned below:

(a) Data input and verification
(b) Data storage and database management
(c) Data transformation

(d) Data output and presentation
(e) Interaction with the user

3.3 Data

Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. The digital map forms the basic data input for GIS. Tabular data related to the map objects can also be attached to the digital data. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organization to maintain their data, to manage spatial data.

3.4 People

GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. The people who use GIS can be broadly classified into two classes. The CAD/GIS operator, whose work is to vectorise the map objects. The use of this vectorised data to perform query, analysis or any other work is the responsibility of a GIS engineer/user.

3.5 Method

And above all a successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization. There are various techniques used for map creation and further usage for any project. The map creation can either be automated raster to vector creator or it can be manually vectorised using the scanned images. The source of these digital maps can be either map prepared by any survey agency or satellite imagery.

4. Data in GIS

A GIS stores information about the world as a collection of themed layers that can be used together. A layer can be anything that contains similar features such as customers, buildings, streets, lakes, or postal codes. This data contains either an explicit geographic reference, such as a latitude and longitude coordinate, or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name. There are two components: spatial data that show where the feature is; and attribute data that provide information about the feature. These are linked by the software. The fact that there are both spatial and attribute data allows the database to be exploited in more ways than a conventional database allows, as GIS provides all the functionality of the DBMS and adds spatial functionality.

4.1 *Spatial data:* Spatial data is spatially referenced data that act as a model of reality. Spatial data represent the geographical location of features for example points, lines, area etc. Spatial data typically include various kinds of maps, ground survey data and remotely sensed imagery and can be represented by points, lines or polygons.

4.2 *Attribute Data*: Attribute data refers to various types of administrative records, census, field sample records and collection of historical records.

Attributes are either the qualitative characteristics of the spatial data or are descriptive information about the geographical location. Attributes are stored in the form of tables, where each column of the table describes one attribute and each row of the table corresponds to a feature.

5. The nature of geographical data

- **Geographical position** (spatial location) of a **spatial object** is presented by 2-, 3- or 4-dimensional coordinates in a geographical reference system (e.g. Latitude and Longitude).

- **Attributes** are descriptive information about specified spatial objects. They often have no direct information about the spatial location but can be linked to spatial objects they describe. Therefore it is often to call attributes "*non-spatial*" or "*aspatial*" information.

- **Spatial relationship** specifies inter-relationship between spatial objects (e.g. direction of object B in relation to object A, distance between object A and B, whether object A is enclosed by object B, etc.).

- **Time** records the time stamp of data acquisition, specifies life of the data, and identifies the locational and attribute change of spatial objects.

5.0 Methods of Data Input in GIS

Data input is the operation of encoding the data and writing them to the database. The creation of clean, digital database is most important and complex task upon which the usefulness of the GIS depends. There are two aspects of GIS database: first, the spatial features and second, the associated attributes representing these features. Spatial data is entered through digitizer or scanners whereas; the attribute data is entered directly through the keyboard.

5.1 Digitizing : Digitizing using a digitizing tablet is also called manual digitizing. The digitizing tablet has a built-in electronic mesh and can sense the position of the cursor and transmit it to the computer. The units of measurement on digitized coverages are in inches. Digitizing begins with a set of tics, which are used later for converting the coverage to real-world coordinates. Two considerations for manual digitizing are: point versus stream mode, and resolution and accuracy.

5.2 Scanning : A scanner converts a paper map to a scanned file, which contains raster data with values of 1 and 0. Pixels with the value 0 represent lines scanned from the paper map, and pixels with the value 1 represent non-inked areas. The scanned file is then converted to a coverage through tracing. GIS packages such as ARC/INFO have algorithms that enable users to perform semi-automatic tracing or manual tracing.

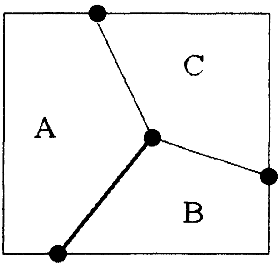
5.3 On-screen Digitizing : On-screen digitizing is an alternative to manual digitizing and scanning for limited digitizing work such as editing or updating an existing coverage. On-screen digitizing is manual digitizing on the computer monitor using a data source such as a DOQ as the background. This is an efficient method for digitizing, for example, new trails or roads that are not on an existing coverage but are on a new DOQ. Likewise, the method can be used for editing a vegetation coverage based on new information from a new DOQ that shows recent clear-cuts or burned areas. Obviously, the major shortcoming of this method is the resolution of the computer monitor, which is much coarser than a digitizing table or a scanner.

6.0 Data Storage in GIS

There are two major types of format in which data can be stored in a GIS. The vector format and the raster format

6.1 Vector Format: In vector models, objects are created by connecting points with straight line (or arcs) and area is defined by sets of lines. Information about points, lines and polygons is encoded and stored as a collection of x, y coordinate. Location of a point feature such as tubewell can be described by a single x, y coordinate. Linear feature such as river can be stored as a collection of point coordinates. Polygon feature, such as river catchment can be stored as a closed loop of coordinates. Vector models are very useful for describing discrete features like data represented by an area. Vector model is not much useful for describing continuously varying features such as soil type.

6.2 Raster Format: Raster format uses regularly spaced grid cells in specific sequence. An element of the grid cell is called a pixel (picture cell). The conventional sequence is row by row from the left to the right and then line-by-line from the top to bottom. Every location is given in two dimensional image coordinates; pixel number and line number, which contain a single value of attributes.



A	A	A	A	A	C	C	C	C	C
A	A	A	A	A	C	C	C	C	C
A	A	A	A	A	A	C	C	C	C
A	A	A	A	A	A	C	C	C	C
A	A	A	A	A	B	B	C	C	C
A	A	A	A	B	B	B	B	B	C
A	A	A	B	B	B	B	B	B	B
A	A	B	B	B	B	B	B	B	B

Fig.1 Vector and Raster Formats

7.0 Questions GIS can answers

Till now GIS has been described in two ways:

1. Through formal definitions, and
2. Through technology's ability to carry out spatial operations, linking data sets together.

However there is another way to describe GIS by listing the type of questions the technology can (or should be able to) answer. Location, Condition, Trends, patterns, Modelling, Aspatial questions, Spatial questions. There are five type of questions that a sophisticated GIS can answer:

Location What is at............?

The first of these questions seeks to find out what exists at a particular location. A location can be described in many ways, using, for example place name, post code, or geographic reference such as longitude/latitude or x/y.

Condition Where is it............?

The second question is the converse of the first and requires spatial data to answer. Instead of identifying what exists at a given location, one may wish to find location(s) where certain conditions are satisfied (e.g., an unforested section of at-least 2000 square meters in size, within 100 meters of road, and with soils suitable for supporting buildings)

Trends What has changed since.............?

The third question might involve both the first two and seeks to find the differences (e.g. in land use or elevation) over time.

Patterns What spatial patterns exists.............?

This question is more sophisticated. One might ask this question to determine whether landslides are mostly occurring near streams. It might be just as important to know how many anomalies there are that do not fit the pattern and where they are located.

Modelling What if................?

"What if..." questions are posed to determine what happens, for example, if a new road is added to a network or if a toxic substance seeps into the local ground water supply. Answering this type of question requires both geographic and other information (as well as specific models). GIS permits spatial operation.

Aspatial Questions

"What's the average number of people working with GIS in each location?" is an aspatial question - the answer to which does not require the stored value of latitude and longitude; nor does it describe where the places are in relation with each other.

Spatial Questions

" How many people work with GIS in the major centres of Delhi" OR " Which centres lie within 10 Kms. of each other? ", OR " What is the shortest route passing through all these centres". These are spatial questions that can only be answered using latitude and longitude data and other information such as the radius of earth. Geographic Information Systems can answer such questions.

7.1 GIS as an Integrating Technology

In the context of these innovations, geographic information systems have served an important role as an integrating technology. Rather than being completely new, GIS have evolved by linking a number of discrete technologies into a whole that is greater than the sum of its parts. GIS have emerged as very powerful technologies because they allow geographers to integrate their data and methods in ways that support traditional forms of geographical analysis, such as map overlay analysis as well as new types of analysis and modeling that are beyond the capability of manual methods. With GIS it is possible to map, model, query, and analyze large quantities of data all held together within a single database.

The importance of GIS as an integrating technology is also evident in its pedigree. The development of GIS has relied on innovations made in many different disciplines: Geography, Cartography, Photogrammetry, Remote Sensing, Surveying, Geodesy, Civil Engineering, Statistics, Computer Science, Operations Research, Artificial Intelligence, Demography, and many other branches of the social sciences, natural sciences, and engineering have all contributed. Indeed, some of the most interesting applications of GIS technology discussed below draw upon this interdisciplinary character and heritage.

9. GIS Use Categories

There are three basic categories of use that GIS can be put to:

- as a spatially referenced database;
- as a visualisation tool; and
- as an analytic tool.

A spatially referenced database allows us to ask questions such as 'what is at this location?', 'where are these features found?', and 'what is near this feature?'. It also allows us to integrate data from a variety of disparate sources. For example to study the dataset on hospitals we might also want to use census data on the population of the areas surrounding each hospital. Census data are published for districts that can be represented in the GIS using polygons as spatial data. As we have the coordinates of the hospitals and the coordinates of the district boundaries we can bring this data together to find out which district each hospital lay in, and then compare the attribute data of the hospitals with the attribute data from the census. We may also want to add other sorts of data to this: for example data on rivers represented by lines; or wells represented by points to give information about water quality. In this way information from many different sources can be brought together and interrelated through the use of location. This ability to integrate is one of the key advantages of GIS.

Once a GIS database has been created, mapping the data it contains is possible almost from the outset. This allows the researcher a completely new ability to explore spatial patterns in the data right from the start of the analysis process. As the maps are on-screen they can be zoomed in on and panned around. Shading schemes and classification methods can be changed, and data added or removed at will. This means that rather than being a product of finished research, the map now becomes an integral part of the research process. New ways of mapping data are also made possible, such as animations, fly-throughs of virtual landscapes, and so on. It is also worth noting the visualization in GIS is not simply about mapping: other forms of output such as graphs and tables are equally valid ways of visualizing data from GIS.

Although visualization may answer some of the questions a researcher has about a dataset, more rigorous investigation is often required. Here again GIS can help. The combined spatial and attribute data model can be used to perform analyses that ask questions such as 'do cases of this disease cluster near each other?' in the case of a single dataset; or 'do cases of this disease cluster around sources of drinking water?' where more than one dataset is brought together. To date, this form of analysis has been well explored using social science approaches to quantitative GIS data. It has not been so well explored using humanities approaches to qualitative data, but this is one area where historians are driving forward the research agenda in GIS.

9.1 GIS Application Areas

GIS are now used extensively in government, business, and research for a wide range of applications including environmental resource analysis, landuse planning, locational analysis, tax appraisal, utility and infrastructure planning, real estate analysis, marketing and demographic analysis, habitat studies, and archaeological analysis.

One of the first major areas of application was in **natural resources management**, including management of

- wildlife habitat,
- wild and scenic rivers,
- recreation resources,
- floodplains,
- wetlands,
- agricultural lands,
- aquifers,
- forests.

One of the largest areas of application has been in **facilities management**. Uses for GIS in this area have included

- locating underground pipes and cables,
- balancing loads in electrical networks,
- planning facility maintenance,
- tracking energy use.

Local, state, and federal governments have found GIS particularly useful in **land management**. GIS has been commonly applied in areas like

- zoning and subdivision planning,
- land acquisition,
- environmental impact policy,
- water quality management,
- maintenance of ownership.

More recent and innovative uses of GIS have used information based on **street-networks**. GIS has been found to be particularly useful in

- address matching,
- location analysis or site selection,
- development of evacuation plans.

The range of applications for GIS is growing as systems become more efficient, more common, and less expensive. Some of the newest applications have taken GIS to unexpected areas. The USGS and the city of Boulder, Colorado have come up with some innovative uses for GIS:

- Global Change and Climate History Project
- Emergency Response Planning

- Site Selection of Water Wells
- Boulder, Colorado, has used GIS to develop a Wildfire Hazard Identification and Mitigation System

9.2 GIS Applications in Agriculture

10.1.1 Precision Farming

Precision Farming (PF) which is otherwise known as Site-Specific farming involves matching resource application and agronomic practices with soil properties and crop requirements as they vary across a site. PF has three requirements such as (i) ability to identify each field location, (ii) ability to capture, interpret and analyze agronomic data at an appropriate scale and frequency, and (iii) ability to adjust input use and farming practices to maximize benefits from each field location. Collectively, these actions are referred to as the "differential" treatment of field variation as opposed to the "uniform" treatment underlying traditional management systems. The result is an improvement in the efficiency and environmental impact of crop production systems.

GIS can integrate all types of information and interface with other decision support tools. It displays analyzed information in maps that allow (a) better understanding of interactions among yield, fertility, pests, weeds and other factors, and (b) decision-making based on such spatial relationships. Many types of GIS software with varying functionality and price are now available. A comprehensive farm GIS contains base maps such as topography, soil type, N, P, K and other nutrient levels, soil moisture, pH, etc. Data on crop rotations, tillage, nutrient and pesticide applications, yields, etc. can also be stored. GIS is useful to create fertility, weed and pest intensity maps, which can then be used for making maps that show recommended application rates of nutrients or pesticides. Variable rate technology in PF is key component which is done for site specific input management. One of the approach for variable rate technology is map-based which includes the steps: grid sampling a field, performing laboratory analyses of the soil samples, generating a site-specific map of the properties through geostatistical technique in GIS environment and finally this map is used to control a variable-rate applicator.

10.1.2 Land capability classification

Land capability is an expression of the effect of physical land conditions, including climate, on the total suitability for use without damage for crops that require regular tillage, for grazing, for woodland and for wild life in short land capability is a measure of suitability of land for use without damage. Irrigability classification is also important in decision making related to land management. Soil irrigability classification is made on the basis of important soil characteristics namely soil texture, depth, available water retention capacity, permeability and alkaline and saline conditions. Land irrigability classification is made taking into consideration, in addition to soil irrigability class, the quantity and quality of water drainage requirements, topography, and economic considerations. The classification for land capability and soil irrigability is commonly done in GIS using all the land and soil parameters required for mapping.

10.1.3 Watershed Management

A watershed is an area from which runoff resulting from precipitation, flows past a single point into a large stream. The watershed approach for natural resource management is an integrated way of looking at land and water resources and their interaction in a given area. Estimation of soil erosion can be done incorporating inputs derived either from conventional techniques or remote sensing into modified universal soil loss equation in GIS environment. Watershed prioritization is done for its management based on severity of degradation.

10.2 GIS Applications in Aquaculture

Planning activities to promote and monitor the growth of aquaculture in individual countries (or larger regions) inherently have a spatial component because of the differences among biophysical and socio-economic characteristics from location to location. Biophysical characteristics may include criteria pertinent to water quality (e.g. temperature, dissolved oxygen, alkalinity: salinity, turbidity, and pollutant concentrations), water quantity (*e.g.* volume and seasonal profiles of availability), soil type (*e.g.* slope, structural suitability, water retention capacity and chemical nature) and climate (*e.g.* rainfall distribution, air temperature, wind speed and relative humidity). Socio-economic characteristics that may be considered in aquaculture development include administrative regulations, competing resource uses, market conditions (e.g. demand for fishery products and accessibility to markets), infra-structure support, and availability of technical expertise. The spatial information needs for decision-makers who evaluate such biophysical and socio-economic characteristics as part of aquaculture planning efforts can be well served by geographical information systems (Kapetsky and Travaglia, 1995).

The first applications of GIS in aquaculture date from the late 1980s (Kapetsky *et al*, 1987). Since then, the use of GIS has been quite limited. Despite this, GIS applications in aquaculture are surprisingly quite diverse, targeting a broad range of species (fish, crustacean, and mollusc) as well as geographical scales, ranging from local areas (Ross, *et al*, 1993), to subnational regions (i.e., individual states/provinces; Aguilar-Manjarrez & Ross, 1995), to national (Salam, 2003) and continental (Aguilar-Manjarrez & Nath, 1998) expanses.

At the present time, the extent of GIS applications in aquaculture include site selection for target species such as fish (Alarcon & Villanueva, 2001), oysters (Chenon *et al.*, 1992), mussels (Scott, Cansado, & Ross, 1998), clams (Arnold *et al.*, 2000), shrimp (Alarcon & Villanueva, 2001), and seaweed (Brown *et al.*, 1999); conflicts and trade-offs among alternate uses of natural resources (Biradar & Abidi, 2000); and consideration of the potential for aquaculture from the perspectives of technical assistance and alleviation of food security problems (Meaden & Kapetsky, 1991; Kapetsky, 1994).A comprehensive listing of works done on GIS application in Aquaculture till date is listed in Table 1. Fisheries and aquatic science professionals are increasingly using geographic information systems (GIS) as a resource management tool. Some case studies on applications of GIS in fisheries and aquatic science research and management are highlighted below.

Case Study 1

A GIS-based model enabling extension personnel, land-use managers, farmers and other interested persons to evaluate potential aquaculture sites in Arizona has been developed in ARCVIEW with the objective to synthesize readily available data into a model capable of predicting locations in which aqua-cultural development could be economically and environmentally viable. Data layers included in the model could be grouped into four categories such as site suitability, water quality, land ownership and infrastructure (Table 2). To test the model's predictive power, existing aquaculture farms were marked on a map generated by the model. Of the 31 farms depicted on the map, 21 occur in areas predicted to have suitable slope and sufficient soil clay content. Of the 10 that occur in areas not predicted as suitable, 5 have the correct slope, 3 have suitable soils and only 2 have neither. Further testing included the generation of species specific (tilapia, trout, bass, catfish and marine shrimp) maps with the locations of existing operations plotted. Of the five models tested, marine shrimp farms were most likely to occur in areas predicted as suitable by the model (67% correct). Trout farms were least likely to have their sites predicted as suitable by the model (27%). Bass, catfish and tilapia farm locations were predicted accurately 65%, 57% and 62% of the time, respectively.

Table 1. GIS applications in aquaculture according to the kind of assessment and scale of study (adapted and updated from, Nath et al, 2000 and Kapetsky and Travaglia, 1995)

Purpose	Geographical region	GIS Software	Author(s)
Large area assessments (*low resolution*)			
Warm water aquaculture	Continental Africa and Madagascar	ERDAS and ARC:INFO	Kapetsky (1994)
Inland aquaculture	Latin America	ARC:INFO	Kapetsky and Nath(1997)
Inland aquaculture	Continental Africa	ARC:INFO	Aguilar-Manjarrez and Nath (1998)
Medium area assessments (*medium resolution*)			
Trout farms	England and Wales	GIMMS	Meaden (1987)
Carp culture (ponds)	Pakistan	Spreadsheet	Ali et al. (1991)
Tilapia and *Clarias* culture in ponds	Ghana	ARC:INFO and ERDAS	Kapetsky et al. (1990a)
Small reservoir fisheries	Zimbabwe	ARC:INFO	Chimowa and Nugent (1993)
Shrimp culture in ponds; fish culture in cages	Johor (State) Malaysia	ERDAS	Kapetsky (1989)
Fish and crayfish farming in ponds	Louisiana (State), USA	ELAS	Kapetsky et al. (1990b)
Pond and cage culture	Tabasco (State), Mexico	IDRISI	Aguilar-Manjarrez (1992)
Fish, shrimp and mollusc culture	Tunisia	ARC:INFO	Ben Mustafa (1994)

Land aquaculture	Sinaloa (State), Mexico	IDRISI	Aguilar-Manjarrez and Ross (1995))
Shellfish and salmon aquaculture	British Columbia(Province), Canada	ARC:INFO	LUCO (1998)
Small reservoir fisheries	Southern Africa	MapInfo, Windisp	ALCOM (1998)
Small area assessments (high resolution)			
Catfish farming	Franklin County, Louisiana, USA	ELAS	Kapetsky et al. (1988)
Shrimp and fish farming in ponds	Gulf of Nicoya, Costa Rica	ELAS	Kapetsky et al. (1987)
Brackishwater aquaculture	Lingayen Gulf, Philippines	SPANS	Paw et al. (1994)
Shellfish culture	Prince Edward Island, Canada	CARIS	Legault (1992)
Salmonid cage culture	Camas Bruaich Ruaidhe Bay, Scotland	OSU-MAP for-the-PC	Ross et al. (1993)
Shellfish culture	Sepetiba Bay, Brazil	IDRISI	Scott et al. (1998)
Shellfish culture	Indian River Lagoon, Florida, USA	ArcView	Arnold et al. (2000)

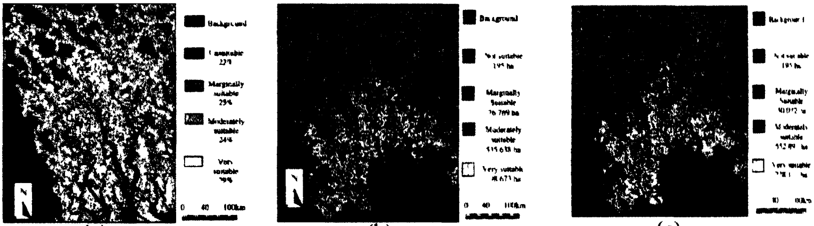
Table 2. Data categories included in the development of the GIS model and the specific data sets comprising each category.

Data Category	Data Used
Site Suitability	Land slope Soil clay content
Water Quality	Temperature Alkalinity PH Total dissolved solids
Infrastructure	Interstates Roads Towns w/population Power lines Railroads
Land Ownership	Private Government Reservation

Case study - II

GIS techniques could be applied to estimate the area of land available for different types of aquaculture development in southwestern Bangladesh, especially the suitability of areas for brackish water aquaculture of the mud crab

(*Scylla serrata*) and the giant tiger shrimp (*Penaeus monodon*) along with satellite remote sensing data (Salam et al , 2003). The models are also used to estimate the production potential and economic characteristics of shrimp and crab farming, and to consider and evaluate alternative land uses in the Khulna region, southwestern Bangladesh. The suitability of available water and land for brackish water shrimp and crab farming in the region is shown in Fig 2.



(a) (b) (c)

Fig 2. Suitability of Water (a), land for shrimp (b) and crab (c) for brackish aquaculture.

Case Study –III

Aguilar-Manjarrez and Ross, 1995 developed a detailed GIS based analytical and predictive tool to guide (shrimp) aquaculture development at a state-level in Mexico. Thirty base layers (thematic maps) were used in the study and included information ranging from pollution sources, population density, general environmental characteristics, land use practices, infrastructure, and water resources (Fig. 3). These layers were organized into 14 criteria, represented either as factors (a measure of the suitability of the criterion relative to the activity under consideration) or constraints (which limits the alternatives under consideration in a binary manner). Two broad categories of factors were identified: physical and environmental characteristics (e.g. water resources, climate, temperature, soils, topography, etc.) and land use type and infrastructure (e.g. agriculture, livestock rearing, population centers, industries, roads, etc.). Constraints used for agriculture and aquaculture were assumed to be identical (e.g. both activities would not be possible in protected land and polluted areas). The identified factors were then grouped into suitability classes. However, it would appear that outcomes from the GIS were more indicative of the true potential for aquaculture because of the range of criteria considered and their integration into suitability models. Moreover, the GIS was able to identify and resolve areas of potential conflict between agriculture and aquaculture (Fig 4). In principle, such information can be helpful to decision makers in terms of exploring alternative land use practices 'prior' to committing them to the landscape.

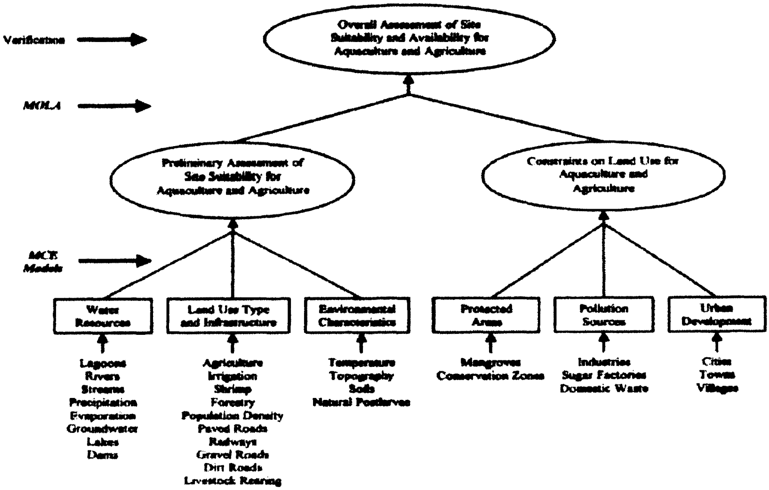


Fig.3 A hierarchical modeling in GIS environment to evaluate suitability of locations for aquaculture and agriculture and resolve associated conflicts, in the Sinaloa state of Mexico (adapted from Aguilar-Manjarrez and Ross, 1995).

Case study –IV

Aguilar-Manjarrez and Nath, 1998 developed a information system for assessing potential of inland aquaculture in Africa. Use of a GIS greatly enhanced the evaluation, especially with regard to the application of objective decision-making methods, quantifying limitations imposed by different production factors, providing estimates of the predicted fish farming potential, and visualizing outcomes. This spatial analysis was limited to assessment of land-based inland aquaculture potential, which for practical purposes implies pond systems. This case study provides a good example of data consolidation from multiple sources. Analytical procedures in the study involved three phases such as (i) criteria identification, classification, and standardization, (ii) integration of primary criteria and (iii) development of models that manipulate and integrate selected criteria.

Criteria used in this study reflects their importance to fish farming, as well as practical considerations of data availability for African countries. The criteria include general environmental characteristics, land use practices, infrastructure, and population distribution data (Fig. 4). Such strategic study gave an important insight for fish farming opportunities 'prior' to encouraging its development. For instance, Zimudzi (1997) reported that after an initial enthusiasm for aquaculture in the 1980s, the majority of Zimbabwean farmers have stopped fish farming due to a variety of reasons including water shortage, poor yields of Nile tilapia due to low water temperatures, and unpredictable survival rates. These same areas were classified as being only marginally suitable in the GIS.

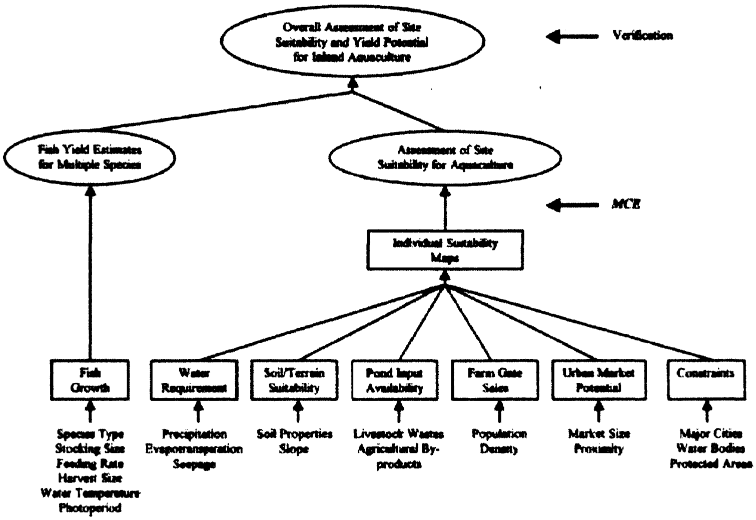


Fig. 4 A hierarchical modeling scheme to assess the suitability of locations, and associated yield potentials for inland aquaculture in Africa (adapted from Aguilar-Manjarrez and Nath, 1998).

11.0 Future Trend

Geographic information systems science and technology, continues to evolve at a rapid pace. Advances are being made with regard to ease of use, manipulation of large (\100 MB) datasets, interoperability of databases among different systems, and in the collection and preprocessing of datasets. Perhaps the most significant development is an increasing trend towards the use of GIS as a component of a larger decision support system. The current release of ArcInfo (Version 8.3, ESRI, Inc.) has replaced the monolithic application model with an object-based component model that facilitates embedding GIS technology in broader application frameworks. This allows developers to produce customized applications utilizing spatial datasets, and has particular usefulness in allowing more robust interaction between models and GIS datasets, and in connecting relational database systems to map displays. Another area in which GIS is playing an increasingly important role is in landscape visualization and 'futuring'. Kapetsky (1998) is of the opinion that this will become increasingly used in both inland fisheries assessments and aquaculture as well.

Technological progress continues to be made in the acquisition of spatially explicit datasets. Increasingly, remote sensed data are becoming available from a rich set of satellite sources and from lower-altitude aerial fly-overs. A substantial commercial industry focused on developing and supplying a broad range of datasets has recently arisen. As previously indicated, the internet has greatly facilitated the distribution of datasets of all types, and most government agencies are beginning to distribute many datasets online. The cost of electronic storage (often less than $20:GB) and increasing capabilities of personal computers is allowing sophisticated GIS analyses to be readily accomplished by inexpensive workstations. Finally, a host of mobile data collection devices, ranging from pen-driven handheld computers to Global Positioning Systems and laser range finders are facilitating more efficient spatial data collection; these will only increase in their sophistication and ease of use.

Given information technology trends, there can be little doubt that future GIS tools will provide a range of functions embedded in various components that can be tailored for specific uses. However, barriers with regard to actually using these tools for real world decision making merit special attention (that has not been forthcoming) and will need to be overcome if GIS is to play an integral role in aquaculture development and management. More specifically, there is the continuing need for appreciation workshops targeted towards decision-makers (i.e. end users) and domain experts to expand knowledge about GIS applications, and the need for relevant coursework in aquaculture and fisheries curricula at the university level (Kapetsky and Travaglia, 1995). Such training and education efforts can go a long way towards making GIS a routine analytical tool in the aquaculture domain. Finally, organizations and individuals must facilitate the migration of GIS tools from the realm of the academic towards an environment within which analysts, subject matter experts and end users can fruitfully collaborate to address issues relevant to the sustainable growth of aquaculture. Work conducted by various governmental agencies in the province of British Columbia (Canada) provides a useful model in this regard, and demonstrates the value of inter- and intra-organizational collaboration with regard to strategic natural resource management initiatives, of which aquaculture is only one component.

12. Conclusions

However, despite some indication that these tools are receiving attention within the aquaculture community, their deployment for spatial decision support in this domain continues to be very slow and only few studies are reported in the literature. This situation is attributable to a number of constraints including a lack of appreciation of the technology of such systems on the part of key decision-makers; limited understanding of GIS principles and associated methodology, inadequate organizational commitment to ensure continuity of these spatial decision support tools and poor levels of interaction among GIS analyst, subject matter specialists, and end users of the technology. However, GIS has several advantages for aquaculture development programs.

References

Aguilar-Manjarrez, J., and S. S. Nath. 1998. A strategic reassessment of fish farming potential in Africa. *CIFA Technical Paper No*. 32. Rome: FAO.

Aguilar-Manjarrez, J., Ross, L.G., 1995. Geographical information systems (GIS), environmental models for aquaculture development in Sinaloa state, Mexico. *Aquaculture Int*. 3, 103–115.

Alarcon, J. F., and M. L. Villanueva. 2001. Using Geographic Information Systems as a site selection tool for aquaculture. In *Aquaculture 2001: Book of Abstracts*. Baton Rouge, LA: World Aquaculture Society.

Arnold, W. S., M. W. White, H. A. Norris, and M. E. Berrigan. 2000. Hard clam (Mercenaria spp.) aquaculture in Florida, USA: Geographic information system applications to lease site selection. *Aquacultural Engineering* 23:203-231.

Biradar, R. S., and S. A. H. Abidi. 2000. Subtle issues in the management of coastal fisheries and aquaculture. In *Subtle issues in coastal management,* 65-75. Dehradun, India: Indian Institute of Remote Sensing (NRSA).

Brown, B., D. W. Keats, J. G. Wakibia, and R. J. Anderson. 1999. Developing community-based seaweed mariculture of Gracilaria gracilis on the South African West coast. In *Conference an Advances on Marine Sciences in Tanzania*, 52. Zanzibar, Tanzania: IMS.

Burrough, P. A. 1990. *Principles of geographical information systems for land resources assessment.* Great Britain: Oxford University Press.

Burrough, P.A., 1986. *Principles of Geographic Information Systems,* 1st ed. Oxford University Press, New York 336 pp.

Chenon, F. H., L. Varet, S. Loubersac, G. Grand, and A. Hauti. 1992. SIGMA, a GIS of the fisheries and aquaculture territorial department. A tool for a better monitoring of public marine ownerships and pearl oyster culture. *Journees Internationales Tenues A Noumea Et A Tahiti,* November 19-24, 1990, 561-570. Noumea, New Caledonia: Orstom.

Kapetsky, J.M., 1994. A strategic assessment of warm-water fish farming potential in Africa. *CIFA Technical Paper*, 27, FAO, Rome.

Kapetsky, J.M., Hill, J.M., Worthy, D.L., 1988. A geographical information system for catfish farming development. *Aquaculture* 68, 311-320.

Kapetsky, J.M., McGregor, L., Nanne, H.E., 1987. A geographical information system and satellite remote sensing to plan for aquaculture development: a FAO-UNEP:GRID cooperative study in Costa Rica. *FAO Fish. Tech. Paper* 287.

Kapetsky, J.M., Travaglia, C., 1995. Geographical information systems and remote sensing: an overview of their present and potential applications in aquaculture. In: Nambiar, K.P.P., Singh, T. (Eds.),

Meaden, G.J., Kapetsky, J.M., 1991. Geographical Information Systems and Remote Sensing in Inland Fisheries and Aquaculture. *FAO Fisheries Technical Paper*, No. 318.

Nath, S. S., J. P. Bolte, L. G. Ross and J. Aguilar-Manjarrez. 2000. Applications of geographical information systems (GIS) for spatial decision support in aquaculture. *Aquaculture Engineering* 23(1-3): 233-278.

Ross, L.G., Mendoza, E.A., Beveridge, M.C.M., 1993. The application of geographical information systems to site selection for coastal aquaculture: an example based on salmonid cage culture. *Aquaculture* 112, 165-178.

Salam M. A, Ross L. G., Beveridge C.M. M 2003, A comparison of development opportunities for crab and shrimp aquaculture in southwestern Bangladesh using GIS modeling, *Aquaculture* 220 (2003) 477-494.

Zimudzi, F., 1997. Locations and Characteristics of Fish Farms in Zimbabwe for Verification Work of a Strategic Re-assessment of Fish Farming Potential in Africa. FAO, Rome.

AN OVERVIEW OF FUNDAMENTAL CONCEPTS AND STATISTICAL TESTS, COMMONLY USED IN QUANTITATIVE RESEARCH

R. N. Subudhi[1]

1. Basic Research-Methodology Concepts

Any standard social science research would generally be comprised of a set of techniques and principles for systemically collecting, recording, analyzing and interpreting data to aid decision makers. So, such research should always be governed by the following principles or guidelines:

a) Attending to the timeliness and relevance of research;
b) Defining the research objectives carefully and clearly;
c) Better not to conduct research for the cases where decisions have already been taken.

We look for gaining or generating some key/ critical information at the end of any research process. Such an information is nothing but the data, that have been collected, analyzed and interpreted by the research scholar/ investigator.

In this paper we shall first enlist and discuss briefly some basic terms as well as types research, which are most commonly in use. This is done just to put the concepts together. Then, in the subsequent section, some basic statistical tools are given to aid the proposed discussion by the author in the workshop.

1.1 Research proposal & research process

A *Research Proposal* is a document that briefly describes the purpose and scope, specific objectives, sample design, data collection procedures, data analysis plan, timetable, and estimated cost for the contemplated project.

The *Research Process* is an interrelated sequence of steps that make up a research project.

The major steps in a *research process* are as follows:

1) Justify the need for marketing research.
2) Define the research objective.
3) Identify data needs.
4) Identify data sources.
5) Choose an appropriate research design and data collection method.
6) Design the research instrument or form.
7) Identify the sample.
8) Collect data, including any relevant secondary data.
9) Analyze and interpret the data.
10) Present the research findings to decision makers.

[1] **Correspondence**: Dr. Rabi Narayan Subudhi, Professor (SRM), KIIT University Bhubaneswar - 751024 Mail: rabisubudhi@gmail.com, rnsubudhi@ksrm.ac.in

2 Types of research

2.1 Basic research versus applied research

Basic research is conducted to generate or create knowledge, while ***Applied Research*** is conducted to solve a problem.

Decision makers can use ***applied research*** to help them develop strategies for their products and services, ranging from fish-breed to fisheries equipments, to even the software, developed by the Institute.

2.2 Qualitative versus Quantitative Research

Qualitative Research is the collection analysis, and interpretation of data that cannot be meaningfully quantified, that is, summarized in the form of numbers. **Quantitative Research** is the collection of data that involves larger, more representative respondent samples and the numerical calculation of results.

2.3 Descriptive Research

The goal of descriptive research, as the name implies, is essentially to describe something. Specifically, it is intended to generate data describing the composition and characteristics of relevant groups of units such as customers, salespeople, organizations and market areas.

2.4 Cross-Sectional Studies are one-time studies involving data collection at a single period in time.

2.5 Exploratory Research versus Conclusive Research

Exploratory Research helps investigators gain some initial insights and may pave the way for further research. **Conclusive Research** helps investigators verify insights and select the appropriate course of action.

Exploratory Research seeks to develop initial insights and to provide direction for any further research needed.

Differences between Exploratory and Conclusive Research can be enlisted as follow:

Research project components	Exploratory research	Conclusive research
a) Research purpose	General: to generate insights about a situation	Specific: to verify insights and aid in selecting a course of action
b) Data needs	Vague	Clear
c) Data sources	Ill defined	Well defined

d) Data collection form	Open-ended, rough	Usually structured
e) Sample	Relatively small; subjectively selected to maximize generalization of insights.	Relatively large; objectively selected to permit generalization of findings.
f) Data collection	Flexible; no set Procedure	Rigid: well laid-out Procedure
g) Data analysis	Informal; typically non-quantitative	Formal; typically quantitative
h) Inferences/ Recommendations	More tentative than final	More final than tentative

3 Methods of data collection

3.1 Focus Group Interviews
In a focus group interview, an objective discussion leader (or moderator) introduce a topic to a group of respondents and directs their discussion of that topic in a non-structured and natural fashion.

3.2 Case Study Method:
The Case study method is an-depth examination of a unit of interest. The unit can be a customer, store, salesperson, firm, market area, website and so on.
By virtue of its insight-generating potential, the case study method is a useful form of exploratory research.

3.3 Direct observation method:
The observational method involves human or mechanical observation of what people actually do or what events take place during a buying or consumption situation. It includes all lab experiments, done in Fisheries and aquaculture sciences.

An Experiment is a procedure in which one (or sometimes more than one) independent (or cause) variable is systematically manipulated and data on the dependent (or effect) variable are gathered, while controlling for other variables that may influence the dependent variable.

A Laboratory Experiment is a research study conducted in a contrived setting in which the effect of all, or nearly all, influential but irrelevant independent variables is kept to a minimum.

A Field Experiment is a research study conducted in a natural setting in which one or more independent variables are manipulated by the experimenter under conditions controlled as carefully as the situation will permit.

3.4 Longitudinal Studies are repeated-measurement studies, that collect data over several periods in time.

3.4 Syndicate Data are secondary data sold by research firms, which we can out source for use in our research. This helps a lot for desk-research or for literature review work.

3.5 Data Warehouse is a centralized database that consolidates company-wide data from a variety of operational systems. ***Data Mining*** is a process of digging deeply into vast amounts of data to extract valuable and statistically valid information that cannot be obtained through queries.

3.6 Questionnaire Format and its types:

Questionnaire Format is an important function of the level of structure and disguise desired, during data collection.

a. Completely Structured Question is one that is presented verbatim to every respondent and with fixed response categories.

b. Completely Non-Structured Question is one that is not necessarily presented in exactly the same wording to every respondent and does not have fixed responses.

c. A Disguised Question is used to examine issues for which direct question may not elicit truthful answers.

4 Validity of survey instrument (Internal & external validity):

Internal Validity is the extent to which observed results are due solely to the experimental manipulation. **External Validity** is the extent to which observed results are likely to hold beyond the experimental setting.

5 Scales and Data Measurement
Measurement is nothing but the assignment of numbers to responses based on a set of guidelines. We may have both metric as well as non-metric data, as the response of the respondents of any research.

Metric Data are data with interval or ratio properties.
Non-Metric Data are data with only nominal or ordinal properties.

5. Scales

We have four types of scales; namely: 1. Nominal, 2. Ordinal, 3. Interval and 4. Ratio scale.

On a ***Nominal Scale***, numbers are no more than labels and are used solely to identify different categories of responses.

An Ordinal Scale is more powerful than a nominal scale in that the numbers possess the property of rank order.

An Interval Scale has all the properties of an ordinal scale, and the differences between scale values can be meaningfully interpreted.

A ***Ratio Scale*** possesses all the properties of an interval scale, and the ratios of numbers on these scales have meaningful interpretations.
We can also have the following types of special-case scales:

5.2 A Graphic Rating Scale presents a continuum, in the form of a straight line, along which a theoretically infinite number of ratings are possible.

5.3 An Itemized Rating Scale has a set of distinct response categories; any suggestion of an attitude continuum underlying the categories is implicit.

5.4 Comparative vs. non-comparative scale

A **Comparative Rating Scale** provides all respondents with a common frame of reference.

A **Non-Comparative Rating Scale** implicitly permits respondents to use any frame of reference, or even none at all.

5.5 Forced-choice & non-forced scales

A **Forced-Choice Scale** does not give respondents the option to express a neutral or middle ground.

A **Non-forced-Choice Scale** gives respondents the option to express a neutral attitude.

5.6 Balanced & un-balanced scales

A **Balanced Scale** has an equal number of positive/favorable and negative/unfavorable response choices.

An Unbalanced Scale has a larger number of response choices on the side of the scale where the overall attitude of the respondent sample is likely to fall.

A **Constant-Sum Scale** has a natural starting point (zero) and asks respondents to allocate a given set of points amongst several attitude objects.

A **Paired-Comparison Rating Scale** consists of a question seeking comparative evaluations of two objects at a time.

A **Single-Item Scale** attempts to measure feelings through just one rating scale.

A **Likert Scale** consists of a series of evaluative statements (or items) concerning an attitude-object. Respondents are asked to rate the object on each statement (or item) using a five-point agree-disagree scale.

A **Semantic-Differential Scale** is similar to the likert scale in that it consists of a series of items to be rated by respondents; however, the items are presented as bipolar adjectival phrases or words that are placed as anchor labels of a seven category scale with no other numerical or verbal labels.

A **Stapel Scale** is a variation of the semantic-differential scale; however, each item consists of just one word or phrase, on which respondents rate the attitude object using a ten-item scale with just numerical labels.

6. Validity and Reliability

Validity is the extent to which a rating scale truly reflects the underlying variable it is attempting to measure.

Content (Face) Validity: represents the extent to which the content of a measurement scale seems to tap all relevant facets of an issue that can influence respondents' attitudes.

Construct Validity assesses the nature of the underlying variable or construct measured by the scale by examining the scales convergent and discriminant validity.

Predictive Validity answers the question "How well does the attitude measure provided by the scale predict some other variable or characteristic it is supposed to influence?"

Reliability measure how consistent or stable the ratings generated by the scale are likely to be.

Test-Retest Reliability measures the stability of ratings over time and relies on administering the scale to the same group of respondents at two different times.

Split-Half Reliability measures the degree of consistency across items within a scale and can only be assessed for multiple-item scales.

B. Tests of Statistical Hypothesis:
Hypothesis is a declaratory statement that is testable. It is a statement about the population that we wish to verify on the basis of available sample information.
It is a provisional answer to the research problem under study that is tested empirically for its validity.

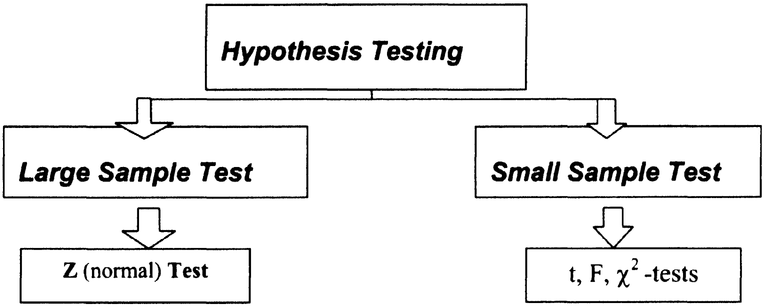
Usually, we test a Null-hypothesis against another alternative hypothesis. The hypothesis, which is tested, is generally denoted as **H_0 (Null Hypothesis)** and the other one, which gives a reverse statement, is denoted as **H_1 (Alternative Hypothesis).**

Criteria for a good Hypothesis:

A good hypothesis should be
- conceptually in clear terms;
- should be testable;
- formulated in specific terms;
- related to available techniques, facts and theories.

Types of Tests of Hypothesis

The following diagram gives us a clear picture on the broad classification of various, commonly used statistical tests:



Sample Testing

Often, we carry out our research on a small sample only (and then we resort to testing a sample for our required information/ result). Result of such a sample test may be, all in probability, true/ correct. But, as our decision (or the result of the sample test) is based on a sample, it may also turn out to be a wrong one.

When we combine our possible decisions on the sample, with the actual state of the population, then we may have four distinct situations. This is well presented in the following table:

Decision from Sample

Actual (true) State		Reject (H_0)	Accept (H_0)
	H_0 True	**Wrong Decision** *1.(Type-I Error) = α	**Correct Decision**
	H_0 False	**Correct Decision**	**Wrong Decision** *2.(Type-II Error) = β

* **Note:** These errors are other wise known as:
(1)Producer's Risk, and
(2)(2) Consumer's Risk, respectively.
Also note that α, the probability of type-1 error, is known as the level of significance of the test (= size of the Critical Region).
Further,
β = Prob {Accepting H_0 | H_1 is true};
where, $(1-\beta)$ is called the power of the test.
It may be mentioned here that, we can't check/ reduce both α and β simultaneously.

The usual practice is to control α at a predetermined low level and subject to this constraint on α (i.e. probability of type-I error), we choose such a test that minimises β (or maximises the power function $1-\beta$). Generally we choose $\alpha= 0.05$ or, 0.01 [i.e. 5% or 1% level of significance].

Test Criterion:

In every test, we generally compute the value of:

$$Z = \frac{|t - E(t)|}{S.E.(t)}$$

Where t is the sample statistic, as discussed earlier.

C. Formulae for sample testing

Large Sample Tests

By "Large Sample" we shall mean that the sample size is very large (usually more than 35) and that the population is NORMAL. Hence all the properties of Normal Distribution are to be applied in such a case. Let us enlist different possible cases and the corresponding test statistics as follow:

Cases:

1. Testing of Mean of a Normal Distribution With Known S. D. (σ):

$$Z = \frac{|X - \mu_0|}{\sigma / \sqrt{n}}$$

2. Testing of mean of a Normal Distribution with unknown S.D.

$$t = \frac{|X - \mu_0|}{S / \sqrt{n}}$$

(σ):
Where,

(S is called sample Standard Deviation).

$$S = \sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}$$

$$Z = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_1}{n_2}}}$$

3. (a) Test of equality of two means with known variances:

(Ho: $\mu_1 = \mu_2$)

(b) Test of equality of two means where variances are not known:

$$t = \frac{|\overline{X}_1 - \overline{X}_2|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where,

$$S = \sqrt{\{\frac{(n_1 - 1)S^2{}_1 + (n_2 - 1)S_2{}^2}{(n_1 + n_2 - 2)}},$$

4. (i) Test of proportions (single sample):

$$Z = \frac{|P_1 - P|}{\sqrt{pq/n}}$$

ii) Test of equality of proportions of two samples:

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

Small Sample Tests

t -Test:

Where can we use this test?

We can apply when:
(i) Sample Size is 30 or less,
(ii) Population variance or standard deviation is unknown.

While testing a hypothesis, the following assumptions are usually made: -

That; (a) the population is normal (or Approximately Normal),
(b) Observations are independently drawn for the random sample,
(c) In case of 2 samples, population variances are assumed to be equal (for the test of equality of Means).

CHI-SQUARE (χ^2) TEST:

Areas of application:
(i) Test of dependence or association between 2 attributes,
(ii) Test of goodness of fit,
(iii) Test of homogeneity (of distributions, correlation Coefficients and population variances).

Test statistic formula is given by:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O-E)^2}{E}$$

This follows a χ^2 distribution, with (n-1) degrees of freedom.

Note: This a reading material, distributed among the trainees and participants of the Training cum National Workshop, conducted by Bio-informatics centre of CIFA and is to be used only as a teaching aid for the proposed discussion by the author in the workshop. This is a compilation work, taken from various standard text books of related area.

REMOTE SENSING AND ITS APPLICATION FOR NATURAL RESOURCE MANAGEMENT

***Rabi N. Sahoo*[1]**

1. Introduction

Over the years, there has been a phase shift of agricultural production and that is from its maximization to optimization. This would be successful only when reliable information related to soil and crop growth condition is available in time to the producer for judicious use of inputs for site specific requirement of the crop. Space technology, especially the data from remote sensing satellites could contribute significantly to improve agriculture. Remote sensing technology being having many advantages like synoptic coverage satellite over large areas, time effectiveness and its practical use in obtaining resources information for rugged an inaccessible terrain, information rich also beyond visible range etc have been proved to be potential tool for generating very profitable agricultural production system. The first remote sensing based pilot project on identification of coconut root-wilt disease in Kerala carried out using infrared aerial photography using an aircraft way back in 1969-70 which was conducted in IARI. Starting with that modest experiment of coconut wilt disease, Indian Remote Sensing Programme has grown into a full-fledged operational programme today. Our remote sensing satellites have become the prime workhorse missions for many developmental applications in the country and is which is more of agriculture-centric. Starting with IRS-1A in 1988, we have launched many remote sensing satellites into orbit such as IRS-1B, IRS-1C, IRS-1D, IRS-P3, Oceansat-1, TES, Resourcesat-1 and Cartosat-1. Today, India has 6 operational IRS satellites operating in orbit making it as one of the largest constellation of remote sensing satellites in the world. The unique combinations of spatial, spectral and temporal resolutions that characterize imaging cameras of IRS satellites have been arrived at, taking into account the sizes of farm holdings, diversity and dynamisms of India's agro-ecosystems.

2.0 Remote Sensing: Definition

Remote sensing is the science of obtaining information about an object or area through the analysis of measurements made at a distance from the object (*i.e.*, not coming in contact with it). In remote sensing, the sensors are not in direct contact with the objects or events being observed. The information needs a physical carrier to travel from the objects/events to the sensors through an intervening medium. The electromagnetic radiation is normally used as an information carrier in remote sensing. The quantity most frequently measured and recorded in images is the electromagnetic energy reflected by the object. The restriction to electromagnetic waves is due to the fact that the observation from a spacecraft excludes other possibilities such as sonic waves, which require a medium like air, water or solid earth for propagation. Other means of indirect observation by for example stationary magnetic or electric fields are not sensitive

[1] **Correspondence:** Rabi Narayan Sahoo, Scientist (Senior Scale) Division of gricultural Physics, Indian Agricultural Research Institute, New Delhi -110 012 Email: rnsahoo@yahoo.com

enough for high geometric resolution measurements. The output of a remote sensing system is usually an image representing the scene being observed. A further step of image analysis and interpretation is required in order to extract useful information from the image. The human visual system is an example of a remote sensing system in this general sense.

2.1.1 Principles

Remote sensing encompasses the set of sensors, platforms, and the data processing techniques that are used to derive information about the physical, chemical and biological properties of the earth surface features (land and ocean) and atmosphere using sensor onboard (air craft or balloons) or space borne (satellite, space shuttles) platforms (**Figure 1**). The oldest form of remote sensing is aerial photography where the sensor system is the camera and film. The field of remote sensing has grown to include electro-optical sensors in optical or microwave regions of electromagnetic spectrum (**Figure 2**) bands which acquire multispectral digital images that can be processed and analyzed by computers. Many of these sensors are on satellites, which regularly orbit the earth.

In optical Remote Sensing, optical sensors detect solar radiation in the visible and near infrared wavelength regions (commonly abbreviated as VNIR) reflected or scattered from the earth, forming images resembling photographs taken by a camera high up in space. Some remote sensing satellites carry passive or active microwave sensors. The active sensors emit pulses of microwave radiation to illuminate the areas to be imaged. Images of the earth surface are formed by measuring the microwave energy scattered by the ground or sea back to the sensors. These satellites carry their own "flashlight" emitting microwaves to illuminate their targets. The images can thus be acquired day and night. Microwaves have an additional advantage as they can penetrate clouds. Images can be acquired even when there are clouds covering the earth surface.

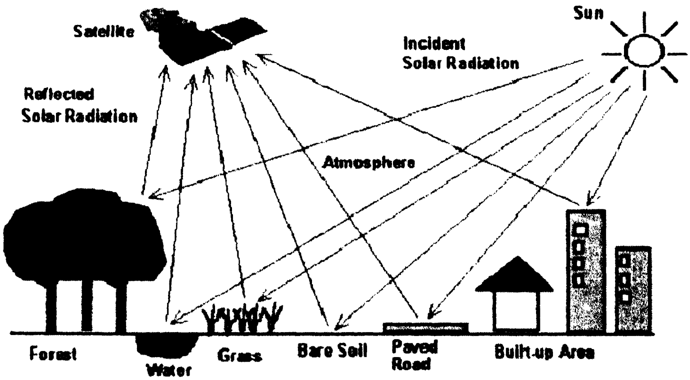


Figure 1. Principle of Remote Sensing

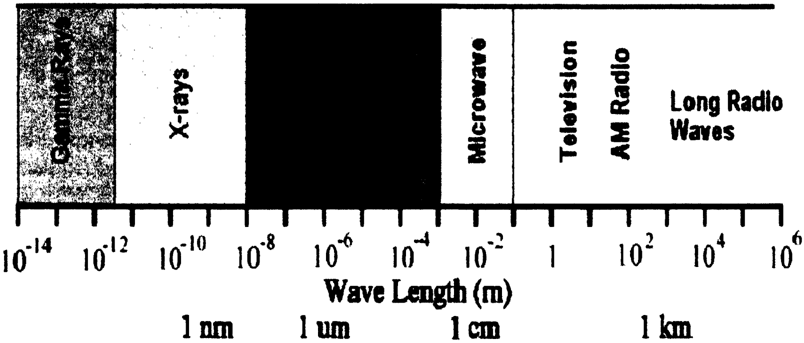


Figure 2. Electromagnetic spectrum showing regions used for remote sensing

A microwave imaging system which can produce high-resolution image of the Earth is the synthetic aperture radar (SAR). The intensity in a SAR image depends on the amount of microwave backscattered by the target and received by the SAR antenna. Since the physical mechanisms responsible for this backscatter is different for microwave, compared to visible/infrared radiation, the interpretation of SAR images requires the knowledge of how microwaves interact with the targets.

In satellite remote sensing of the earth, the sensors are looking through a layer of atmosphere separating the sensors from the Earth's surface being observed. Hence, it is essential to understand the effects of atmosphere on the electromagnetic radiation traveling from the Earth to the sensor through the atmosphere. The atmospheric constituents cause wavelength dependent absorption and scattering of radiation. These effects degrade the quality of images. Some of the atmospheric effects can be corrected before the images are subjected to further analysis and interpretation.

A consequence of atmospheric absorption is that certain wavelength bands in the electromagnetic spectrum are strongly absorbed and effectively blocked by the atmosphere. The wavelength regions in the electromagnetic spectrum usable for remote sensing are determined by their ability to penetrate atmosphere, which depends on its transmittance properties. These regions are known as the atmospheric windows. Remote sensing systems are often designed to operate within one or more of the atmospheric windows in optical and microwave regions of electromagnetic spectrum. These windows exist in the microwave region, some wavelength bands in the infrared, the entire visible region and part of the near ultraviolet regions (**Figure 3**). Although the atmosphere is practically transparent to x-rays and gamma rays, these radiations are not normally used in remote sensing of the earth.

When solar radiation hits a target surface, it may be transmitted, absorbed or reflected. Many remote sensing systems operate in the wavelength region in which the reflected energy predominates. The reflectance properties of the earth surface features may be quantified by measuring the fraction of incident energy that is reflected. This is measured as a function of wavelength and is called spectral reflectance curve. The configuration of spectral reflectance curve gives insight into the spectral characteristics of an object based on which it is identified.

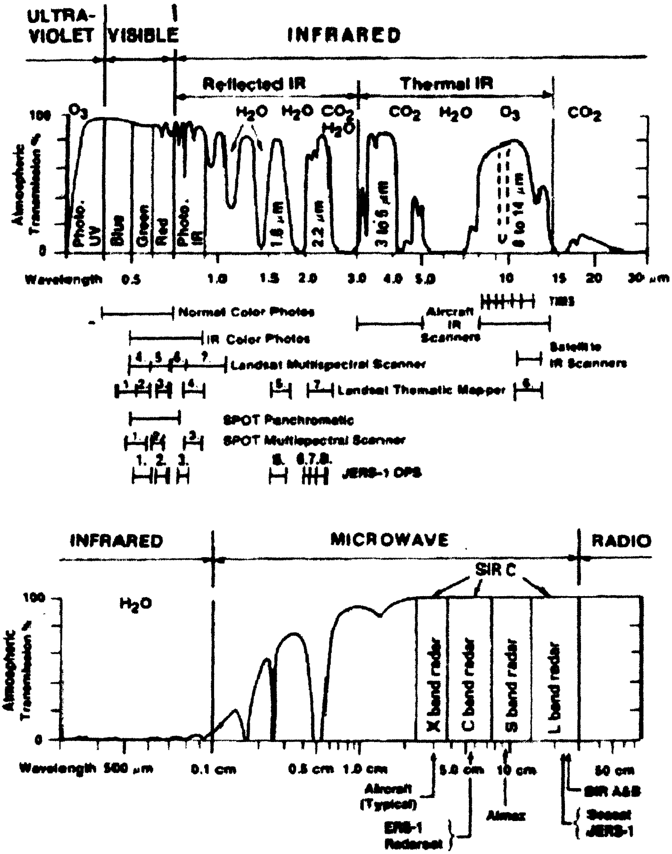


Figure 3. Diagram showing Atmospheric transmission (windows) in visible, infrared and microwave regions and wavelength bands used for different remote sensing systems (Source: Sabins, 1997)

The typical spectral reflectance curves of soil, vegetation and water features are given in the **Figure 4**. The spectral reflectance curve for healthy green vegetation manifests the 'peak and valley' configuration. The soil curve shows considerably less 'peak and valley' variation in reflectance. The most distinctive characteristic of water is the energy absorption at near infra red wavelengths. Though these broad feature types are normally spectrally separable, the degree of separation differs in different wavelength regions. For example, water and vegetation might reflect nearly equally in visible wavelengths, yet these features are almost always separable in near infrared wavelengths. The spectral reflectance curve serves as a unique signature of the feature. In principle, a feature can be identified from its spectral reflectance signature if the sensing system has sufficient spectral resolution to distinguish its spectrum from those of other materials. This premise provides the basis for multispectral remote sensing. Even within a given feature type, the proportion of reflected, absorbed and transmitted energy will vary at different wavelengths. Thus two features may be distinguishable in one spectral range and be very different in another wavelength band.

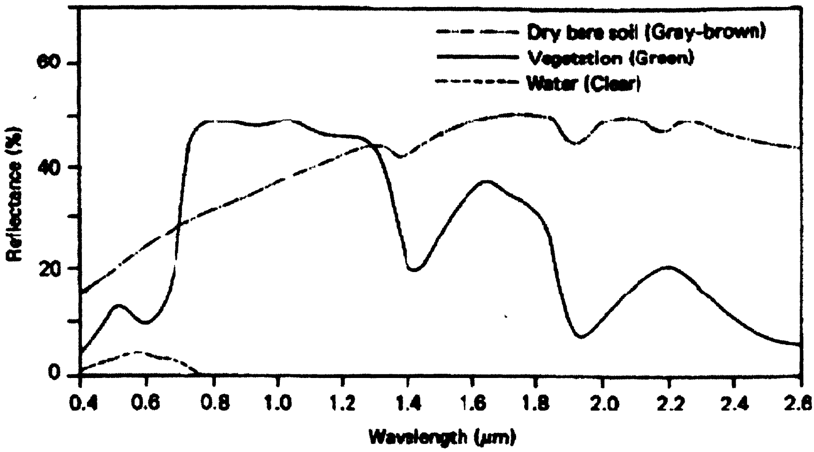
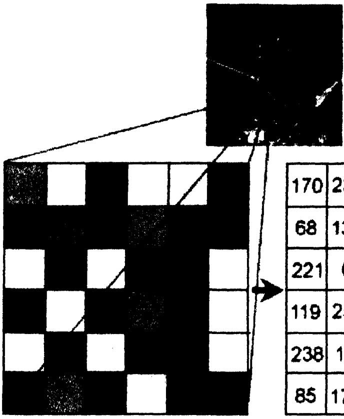


Figure 4 Typical spectral reflectance of soil, vegetation and water.

Remote sensing images are representations of parts of the earth surface as seen from space. The images may be analog or digital. Aerial photographs are examples of analog images while satellite images acquired using electronic sensors are examples of digital images. A digital image comprises of a two dimensional array of individual picture elements called pixels arranged in columns and rows (**Figure 5**). Each pixel represents an area on the Earth's surface. A pixel has an intensity value and a location address in the two dimensional image. The intensity value represents the measured physical quantity such as the solar radiance in a given wavelength band reflected from the ground, emitted infrared radiation or backscattered radar intensity. This value is normally the average value for the whole ground area covered by the pixel.



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

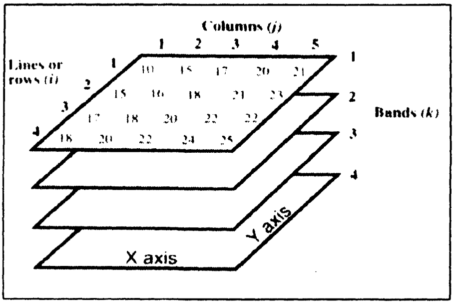


Figure 5. A remote sensing image showing pixels with intensity value and location address

2.1.2 Resolution of Remote Sensing Data

The information acquired through remote sensing imaging is resolution dependent. Several types of resolutions are considered in remote sensing studies those are (i) Spatial resolution (ii) Spectral resolution (iii) radiometric resolution and (iv) temporal resolution.

Spatial resolution refers to the size of the smallest object that can be resolved on the ground. In other words, it is the minimum distance between two objects that a sensor can record distinctly. In a digital image, the resolution is limited by the pixel size, i.e. the smallest resolvable object cannot be smaller than the pixel size. The intrinsic resolution of an imaging system is determined primarily by the instantaneous field of view (IFOV) of the sensor, which is a measure of the ground area viewed by a single detector element in a given instant in time. However, an IFOV value is not in all cases a true indication of the size of the smallest object that can be detected. An object sufficient contrast with respect to its background can change the overall radiance of the given pixel so that the object becomes detectable. Spatial resolution of a remote sensing system must be appropriate if one is to discern and analyze the phenomenon of interest. To move from detection to identification, the spatial resolution must improve by about 3 times. To pass from identification to analysis a further improvement in spatial resolution of 10 or more times may be needed.

A **"High Resolution"** image refers to one with a small resolution size. Fine details can be seen in a high resolution image. On the other hand, a **"Low Resolution"** image is one with a large resolution size, i.e. only coarse features can be observed in the image.

Spectral resolution is determined by the band widths of the channels used in the imaging system. High spectral resolution is achieved by band widths which collectively are likely to provide more accurate spectral signature for discrete objects than by broad bandwidths. Spectral resolution varies from a single band panchromatic system, four or seven multispectral band system in IRS SPOT and many satellite systems to many hyper-spectral bands of TERRA or AQUA MODIS satellite system.

Radiometric Resolution refers to the smallest change in intensity level that can be detected by the sensing system. The intrinsic radiometric resolution of a sensing system depends on the signal to noise ratio of the detector. In a digital image, the radiometric resolution is limited by the number of discrete quantization levels used to digitize the continuous intensity value. With a given spectral resolution, increasing the number of quantizing levels or improving the radiometric resolution will improve discrimination between scene objects. Interdependency between spatial, spectral and radiometric resolutions for each remote sensing system affect the various compromises and trade offs.

Temporal resolution is an important consideration when determining the resolution characteristics of a sensor system. It is defined by the repetitive period of the sensors. This is very important to monitor any temporal dynamics of the features. For example, temporal growth profile of a crop monitored through remote sensing derived parameters helps in identifying and discriminating it.

2.1.3 Remote Sensing as a Data Source

As Remote sensing is the noninvasive and non destructive gathering of information about the features, it has become a valuable source of input for GIS databases because remote sensing data (in particular satellite-borne) provides synoptic viewing, data comparability, repeat coverage, the capability for historical record, as well as the provision of data in the non-visible part of the spectrum.

Again satellite systems such as SPOT, NOAA -AVHRR and IRS -WiFS acquire data for large areas in a short time period, there by providing essentially uniform coverage with respect to data and level of detail. Such data are already in digital form and are provided more or less standard formats. These data products are available for almost all the earth's land areas and are inexpensive relative to alternate sources. Although imageries are not planimetrically correct, pre processing can often bring data to acceptable levels of geometric accuracy with only modest effort.

There are five main sources of information that can be exploited by remote sensing systems. These relate to variations in the recorded signal as a function of (i) wavelength (spectral), (ii) angle (directional) (iii) wave polarization (iv) location (spatial) and (v) time (temporal). These are discussed as below.

There are two main ways in which the relationship between surface properties and spectral response can be exploited. At one level, the aim is to discriminate different type of surface materials. In this case, the objective is to identify those wavelengths at which the contrast between their reflectance, emittance, or scattering characteristics is maximized. Since all surface materials can be differentiated at a given wavelength, it is common to record data in several parts of the electromagnetic spectrum (*i.e.* multispectral remote sensing). A subsequent aim may be to identify the nature of the surface material, i.e. to assign each a label from a set of pre-defined classes, typically expressed in terms of land cover. The second major use of multispectral data is to estimate values for selected properties of the observed surface materials such as to derive information on above ground biomass and Leaf Area Index (LAI) etc of vegetation. This is commonly based on linear combinations of data recorded in two or more spectral bands preferably in red and Near infra-red wavelengths which is otherwise referred as vegetation indices. The enduring attraction of vegetation indices despite their well-known limitations lies in their conceptual and computational simplicity. For example Normalized Difference Vegetation Index (NDVI) has been received enduring popularity for mapping and monitoring vegetation at regional and global scales (http://free.vgt.vito.be). Attempts have been made to develop physically based models to account for the optical properties of the leaves for their biophysical properties (Jacquemoud and Baret, 1990). These models are less data dependent and site specific and can be inverted to derive biophysical propertied of the vegetation (Goel, 1989).

The detected reflectance of most surface features varies as a function of the angles at which they are illuminated by the sun and received by the sensor. The angular distribution of the reflected radiation is described by the bidirectional reflectance distribution function (BRDF). Various mathematical models have developed and used with BRDF for estimation of biophysical properties of vegetation and soil through model inversion techniques (Goel 1989).

The amount of radiation reflected, emitted or scattered from the earth surface varies spatially in response to changes in the nature (type and properties of the materials. These variation may be continuous, discrete, linear or localized depending on the controlling environmental processes. They may be manifest at a variety of different spatial scales. The relation between surface type, surface properties and spatial variability in land leaving radiance has been exploited using measures of

- Texture - the statistical variability of the detected signal, typically based on the grey level co-occurrence matrix, measured at the level of individual pixels
- Pattern - including the size and shape of discrete spatial entities (regions), typically land cover parcels, identified within the scene as well as the spatial relations between them
- Context - referring to the structural and semantic relations between discrete spatial entities identified within the scene

- Temporal variation in reflectance properties may be related to the diurnal or/and seasonal (*eg.* phenology) effects, episodic events (*e.g.* rainfall or fire), anthropogenic influences (*e.g.* deforestation or long term climate change.

As one possible source of GIS input data, remote sensing is useful in four capacities i.e. (i) as a display backdrop (ii) for generation of thematic maps (iii) for derivation of input variables for models and (iv) as a real time link.

3. Applications of Remote Sensing in Agriculture

For better crop production, timely information on soil resources mainly and its fertility and limitations, different parameters governing crop growth condition (such as biotic and abiotic stresses), disasters like drought and flood is very required. Remote sensing provides such valuable information on spatio-temporal scale for which any other alternative means are found to be highly expensive and sometimes not affordable. This technology holds great promise for site specific crop management because of its potential for monitoring spatial variability over time. The different components of agricultural resources which can be monitored and assessed by remote sensing techniques are listed in Table 1.

Table 1. Possible Uses of Remote Sensing Techniques in Assessment and monitoring of Land, Water and Crop Resources.

1.	Soil and land Resources	Soil characteristics, soil survey, land use/land cover change, land use planning, and assessing different soil parameters, assessment of soil and land productivity.
2.	Degraded and wasteland monitoring and management, soil erosion and soil conservation	Degraded and wasteland mapping and monitoring extent of saline and sodic areas and effect of reclamation measures: seasonal water logging, extent of wetland areas. Fluvial erosion, extent and depth of gullies; eolian erosion, sand dune movement: shifting cultivation: desertifiction, arid land forms
3.	Integrated watershed management	Watershed characterization and prioritization, generation of action plans, impact assessment of soils and water conservation measures.
4.	Agricultural disasters	Agricultural drought monitoring, extent and damage due to floods
5.	Water resources	Surface water bodies, ground waster targeting; soil moisture status and soil moisture mapping: irrigation scheduling evapotranspiration estimation
6.	Crop resources	Crop identification discrimination & acreage estimation: crop condition assessment, crop stress detection (water and diseases): different parameter retrieval, crop yield modeling and crop production forecasting

3.1 Soil and Land Resources Inventory

Remote sensing technologies through air borne and satellite-based sensors are important tools for efficient mapping as well as monitoring of soil resources. It has been demonstrated that space borne data afford greater degree of accuracy and is time saving for mapping. Soil surveyor detects different soil formative environments through visual interpretation of geological maps, topographical maps and satellite images. The spatial extents of the soil formative environments are then used to delineate soil-landscape units known as physiographic units. Physiographic units are based on the relationships between these environmental conditions and the soil mapping units. These physiographic units are often initially delineated on a set of air photos /satellite images, then field checked and compiled onto a base map. Soil profiles were then studied in these physiographic units to characterize and classify the soils to prepare physiographic-soil map. The use of satellite data for soil survey is either by digital analysis of images or by the visual interpretation of FCCs (Hiwig and Karale, 1973; Hilwig, 1975). The digital approach of image interpretation is quantitatively oriented, using automated classification method, pixel by pixel. Careful visual image interpretation can be better than digital image classification for soil resource mapping. Ancillary data used with satellite images to improve their visual interpretability.

The recent technology on hyperspectral remote sensing helps in quantitative assessment of different soil parameters thereby increasing the efficiency of expensive and time-consuming soil-related studies. The rapid nature of the measurement (through this technique) allows soil variability to be more adequately sampled than with conventional approaches and thereby facilitates risk-based approaches to soil assessments. A study was conducted in IARI for for quantitative assessment of 16 physico-chemical parameters such as texture, moisture, colour, organic matter, soil reaction, available N, P and K, exchangeable Ca, Mg, Fe, Mn, Cu and Zn, Calcium carbonate and oxide contents through the hyperspectral remote sensing techniques. For site specific management, homogenous land units can be developed using both remote sensing and GIS techniques and fertilizer requirement for each homogenous land units or fertility units can be recommended for various crop which has been successfully done for Karnal District by IARI.

There are various land evaluation methods used for assessing the potential and productivity of soil for agricultural purposes (as given below). Now a days remote sensing tool have been used for assessing productivity potential of the land for agriculture purpose.

i). Land Capability Classification

ii) Soil and Land Irrigability Classification

iii) Parametric methods of Land Evaluation

iv) Land Productivity Index (Storie, 1976)

v) Soil Productivity Index (Ricquier et al. 1970)

vi) FAO Framework of Land Evaluation (1976)

3.2 Expanding the Scope for Agriculture

On expanding the scope for agricultural activities in the wastelands, you are aware that Government had set up in 90s, a High Level Committee with Shri Mohan Dhariya in Chair. The Committee recommended that (i) all the culturable and environmentally degraded wastelands are to be made productive, (ii) generate locally the employment opportunities in this process to the millions of rural poor, (iii) raise food grain production from 210 to 450 MT with such endeavours, and (iv) bring 30 percent of the areas, especially environmentally degraded lands, under green cover. To implement such recommendations and related interventions on the ground, Ministry of Rural Development was in need of knowing the extent of degraded areas, their locations in different States, Districts, Villages and in Watersheds. Department of Space, responded to the Ministry, carried out nationwide wasteland mapping project based on satellite images and reported the wastelands of around 64 Mha in 1998.

Since then, Ministry of Rural Development implemented several wasteland development programmes and was keen to see the impact on the ground. Accordingly, Ministry sponsored another project to Department of Space - called 'National Wastelands Updation Mission' project. Using IRS data of year 2003-05, it was found that the wastelands have come down to around 56 Mha. It shows clearly that in 6 to 7 years there has been reduction, to the extent of more than 8 Mha, in the wastelands. In three States namely Meghalaya, Nagaland and Gujarat the reduction is more than 10 percent. It is important to highlight that reduction in wastelands was mainly a gain to the agricultural and social forestry activities.

The National Wastelands Atlas has brought out the fact that about 35 Mha of wasteland can be reclaimed and made suitable for cultivation, especially agro-horticulture, coarse gains, bio-fuel plantation etc., with appropriate interventions.

3.3 Degraded land Management and restoration of Land Capability

In the Indo-Gangetic plains of Uttar Pradesh, historically the most fertile and irrigated agro-ecosystems, large areas turned infertile. It was primarily due to lack of the proper drainage associated with shallow ground water table. As a result, about 0.6 Mha geographical areas got severely affected by water logging, soil salinity and alkalinity. In order to control soil degradation and improve agricultural productivity a World Bank funded 'U.P. Sodic Land Reclamation Project' was undertaken and executed by the U.P. Bhumi Sudhar Nigam supported by Remote Sensing Applications Centre - Uttar Pradesh (RSAC-UP), Lucknow, which provided satellite based information. The wastelands maps, prepared as a part of the national wasteland mapping mission, were used to identify areas with various levels of severity; 10 districts having maximum area of sodic soils were selected from these maps. Further, based on the ground water quality reported from ground observations, villages were selected for reclamation. For actual execution and identification of plots within the villages, sodic lands were mapped on cadastral scale (1:4,000) using aerial photography for about 900 villages. Specific physical intervention like addition of gypsum & improvements of drainage networks as well as concurrent monitoring and evaluation contributed to the success of this project in reclaiming the degraded soils and restoring the productivity.

The soils after reclamation are producing cereals to the extent of 2.6 Tonnes per hectare and increasing family income by 50 percent, against virtually nothing before the treatment. The UP Sodic Land Reclamation Project has estimated the net returns of Rs 700 crores from an area of around 36000 ha. This indicated that for 1 Mha salt-affected lands, the net returns would be nearly Rs 2000 crores. The project has helped in poverty reduction, increased family income in about 2,28,000 families belonging to small and marginal farmers and more importantly, effecting ecological restoration of sodic soils. It is worth highlighting that remote sensing and GIS inputs, though costing just 2 percent of the project cost, contributes significantly to the success of project.

There are 7 Mha saline and alkaline soils in the country as per National Wasteland maps, which need treatment. The wasteland maps are the only spatial information, which could serve the purpose of making specific physical interventions towards ecological restoration and improving productivity. It is important that the success of 'U.P. Sodic Land Reclamation Project' could be replicated in these areas.

3.4 Watershed characterization and prioritization

With synoptic, repetitive and wide coverage in narrow and discrete bands satellite remote sensing has been extremely valuable tool for characterization of watershed for soil conservation measures to be adopted for sustainable development. Identification of drainage and stream network, land cover/land use, surface water bodies, land forms etc in a watershed has been made possible using satellite data. Work on characterization of several watersheds (in Narmada basin, Ukai catchment)and correlation of annual runoff with land cover, development of erosion indices for prioritization of watershed for soil conservation measures has been reported by Sahai (1989).

Under the Integrated Mission for sustainable development (IMSD) programme of Department Space, studies are being carried out using remote sensing, GIS and other conventional data for generation of action plan packages on watershed/tehsil block basis in 175 districts all over India. In this programme, a natural resources data base e.g. hydrogeomorphology, water quality and depth, soils, vegetation land use/land cover, drainage and watershed, slope, transport and settlements etc. are created using remotely sensed and other conventional data. Then integration of natural resources database, social-economic database and contemporary technology is carried out in GIS environment and action plans are generated for development of land and water resources such as soil conversation, ground water exploration an recharge, surface water harvesting, alternate land use practices and use of waste land into productive use. These action plans are provided to the district and state administration for implementation. The integration of remotely sensed satellite date with ground survey in planning and development of watersheds holds great promise.

3.5 Crop Discrimination and Cropping System Analysis

Diversification of agriculture is a lesson learnt from the green revolution. To stabilize the productivity levels, diversification holds the key. Diversification of agriculture calls for ecological zonation in terms of uniqueness of soil types, crops, climate and topography. IARI has carried out cropping system analysis for

Trans Gangetic Plain of the country using time series satellite data available in public domain. New techniques were developed to discriminate different crops, cropping pattern using these products. Even if the potential of this data was found to discriminate early, timely and late sown crop in a region. The cropping system analysis has provided valuable inputs for crop diversification and intensification planning. Same work was also carried out by Dept of Space form whole state Punjab and West Bengal. While implementing the same, long-term sustainability and cropping system performance as well as feasibility of applying the precision farming concept need to be appropriately addressed. The cropping system analysis needs to be extended to national level to maximize the agricultural productivity.

3.6 Crop Parameter Retrieval and Prioritization for Intervention

Understanding and predicting crop phenology and canopy development is important for many reasons including improving the efficacy of management practices and increasing accuracy of simulation models and decision support systems. Biophysical parameters such as leaf area index (LAI), fraction of absorbed photosynthetic active radiation (fAPAR), fraction vegetation cover (FVC), chlorophyll content, canopy height, roughness etc. have been identified as the most important physical properties of the terrestrial surface due to their specific roles in biosphere-atmosphere interactions, and they play unique role in global climate change studies. The knowledge of canopy biophysical variables and agrometeorology is of prime interest in many applications and can be used as inputs to crop growth and yield simulation models. The role of remote sensing in retrieval of some of these parameters could be explained in terms of parameters that can be obtained from Earth observing satellite data. The parameters of interest amenable to remote sensing based assessment are grouped under three groups: (i) crop phenology (emergence, greening, peak vegetative stage, senescence, length of growing season); (ii) biophysical parameters (LAI, fAPAR, FVC, chlorophyll content, canopy roughness etc.); and (iii) Agrometeorological parameters (Albedo, surface temperature, soil moisture etc.). IARI experiment has been successfully conducted to retrieve Leaf Area Index (LAI), chlorophyll (Cab) and moisture content (Cw) of wheat crop at regional scale in Trans Gangetic Plain from free available MODIS data products using physical process based radiative transfer model and developed a composite crop health index based on LAI, chlorophyll and moisture content of the crop for better site specific crop management practices.

3.7 Water and Nutrient Management

Nutrient and water stress management is another area where remote sensing data has been potential used in country and abroad. Detecting nutrient stresses using remote sensing and combining data in a GIS can help in site-specific applications of fertilizers and soil amendments such as lime, manure, compost, gypsum, and sulfur, which in turn would increase fertilizer use efficiency and reduce nutrient losses. In semi-arid and arid tropics, precision technologies can help growers in scheduling irrigation more profitably by varying the timing, amounts and placement of water. For example, drip irrigation, coupled with information from remotely sensed stress conditions (e.g., canopy-air temperature difference), can increase the effective use of applied water thereby reducing runoff and deep percolation.

Pests and diseases cause huge losses to crops. If remote sensing can help in detecting small problem areas caused by pathogens, timing of applications of fungicides can be optimized. Recent studies show that pre-visual crop stress or incipient crop damage can be detected using radio-controlled aircraft and near-infrared narrow-band sensors. Likewise, airborne video data and GIS have been shown to effectively detect and map black fly infestations in citrus orchards, making it possible to achieve precision in pest control (Everitt *et al.*, 1994). Perennial weeds, which are usually position-specific (Wilson and Scott, 1982) and grow in concentrated areas, are also a major problem in precision agriculture. Remote sensing combined with GIS and GPS can help in site-specific weed management.

3.8 Water Resources and Management

Increased interest in the utility of remote sensing techniques for monitoring waster resources and its efficient management in agriculture has necessitated study of the possibility for its use in assessing surface water bodies and ground water potential, evaluation of soil moisture status and its mapping, irrigation scheduling for crops and estimating evapotranspiration for better utilization of water resources. The potential of microwave remote sensing in measurement of soil moisture and monitoring land resources has been demonstrated.

The launching of European Remote Sensing Satellites ERS-1 and ERS-2 with Synthetic Aperture radar (SAR) and Canadian Satellite RADARSAT has opened up new vistas in utilization of this frontier technology in studying dynamics of diversified soil and crop growth processes.

3.8.1 Irrigation scheduling and evapotranspiration

Experiments based on satellite imageries and ground based observations provide evidence that the remote sensing technology can be profitably used as an aid in scheduling irrigation and estimating evapotranspiration or water use by crops. IARI through its long term filed experiments has potentiality demonstrated irrigation scheduling based on remotely sensed canopy temperature to attain higher yield and water use efficiency of crops.

3.9 Crop Forecasting

In the globalised world, agricultural statistics provides the vital informatics support. There are hundreds of crucial decisions at different levels, which are taken purely based on agricultural statistics. To strengthen this, you are aware that Ministry of Agriculture supports a remote sensing based the nationwide project called pre-harvest Crop Acreage and Production Estimation (CAPE), providing in-season crop statistics with reasonable accuracy. IARI has been the prime mover in this concept and many of IARI's students who are with ISRO have contributed to this unique Project. Based on the experience of CAPE, synthesizing state-of-the-art in econometrics, agricultural meteorology and remote sensing, Forecasting Agricultural Output using Space, Agro-meteorology and Land-based Observations (FASAL) has been embarked upon for making multiple nationwide forecasting. It has been one of the operational applications in remote sensing providing the rabi wheat forecast reasonably with high accuracy much before the harvest.

FASAL project, for the cropping season 2004-05, made the national wheat forecast 72 MT in the month of February-March 2005 itself. The final estimate of 72.9 MT was made by Ministry of Agriculture some times in the month of June-July. Timeliness, along with spatial attributes, is the most critical component for policy decisions. These efforts have thus brought down the gaps between crop production and post-harvest technology, pricing and policy decisions. We have also been successful in integrating remote sensing technology in the priority areas such as horticulture, crop diversification, land degradation mapping and so on.

4. Monitoring Agricultural Disasters

4.1 Managing Flood Plain Agro-ecosystem

Strength of remote sensing and GIS lies in reaching out and unfolding the various linkages and the underlying factors existing between state of the natural resources and environment and incidence of disaster induced low productivity and food insecurity. The advantage of remote sensing technology has been its inherent capability that provides synoptic and repetitive high resolution imageries over large areas and thus maps out the spatio-temporal variability of floods and crops. Geographic Information System (GIS) establishes dynamic linkages between biophysical factors with socio-economic variables for analyzing low productivity through modeling and simulation.

4.1.1 Targeting the vulnerability

Vulnerability of agriculture to flood is well known. Remote sensing captures the vulnerability through near real time flood monitoring, its impact assessment and forecast modeling. Mapping of flood-inundated areas is one of the most successful applications of satellite remote sensing and this is routinely operational in the country since 1988. Because of the unique spectral signature, it is possible to map areas under standing water, areas from where flood water receded, submerged standing crop areas, sand casting of agricultural lands, breaches in the embankments, marooned villages and towns etc. Using multi-date satellite imageries, the extent of damage due to crop loss, destruction of infrastructure facilities etc can be assessed. Satellite remote sensing and GIS techniques have been integrated (Venkantachary *et al,* 2001) in Brahmaputra river basin to provide information on flooded area and damage to cropping lands, roads and rail tracks. Global Positioning System (GPS) is being used to aid in the development of Digital Elevation Model (DEM) for a flood prone area in Andhra Pradesh, to enable assessment of spatial inúndation at different water levels in the river. When satellite derived land use and land cover and ancillary ground based socio-economic data is assessed to provide location specific flood warnings. Remote sensing data are evaluated for integration with existing forecasting models. Also microwave data from RADARSAT is used in conjunction with optical data to overcome the limitation of cloud cover. A proto-type system consisted of comprehensive database design, space-based near real time monitoring tools, modeling framework, networking and user interface was synthesized by ISRO in 1998. The date wise flood inundation and recession pattern during the event year 2003 and its spatial extent during 1988 to 2003 in Eastern Gangetic Flood Plain of Bihar is given in **Figure** 6. The year 2003 was flooding event year for Ganges and peak flooding period was from September 15 to October 15, 2003. Flood maps for different dates of 2003 and different years derived from various satellites such as MODIS, LANDSAT (Source: NASA supported Dartmouth Flood Observation Center, USA) and other microwave sensors were overlaid to obtain the spatio-temporal dynamics of flood. Vector layer of location map and digital elevation map was also overlaid over the composite flood map.

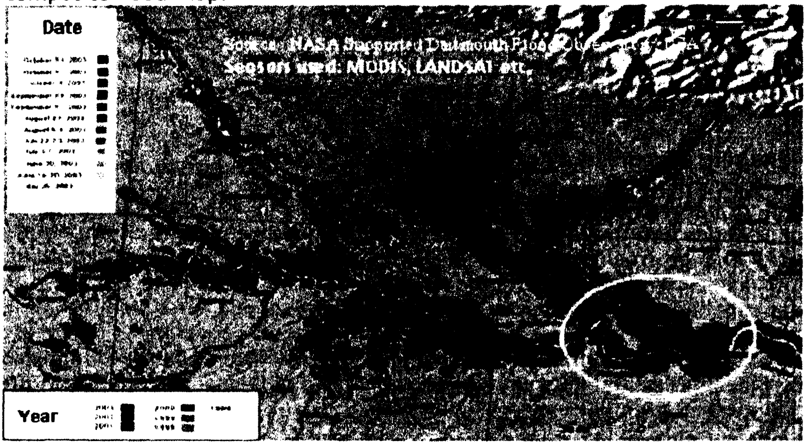
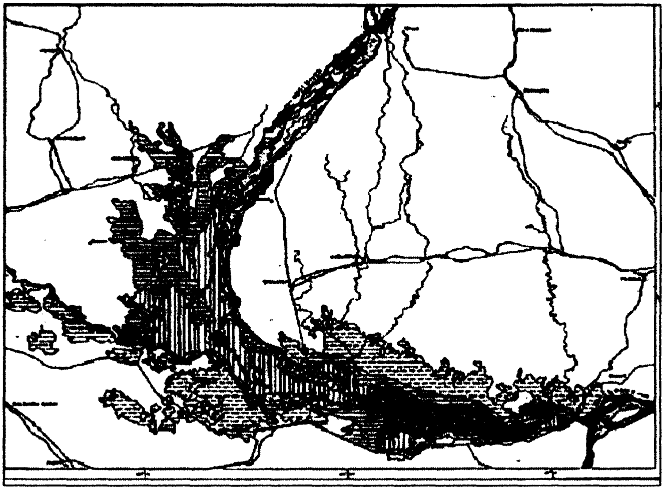


Figure 6. Spatio-temporal dynamics of flood during 2003 and for the period 1998-2003.

Flood risk zoning is an essential tool for appropriate land use planning in flood prone areas. It aims at demarcating areas likely to be affected by floods of different magnitude, probability levels, and risks associated for living and economic activities, which in turn specifies the types of permissible developments in these zones to be undertaken. Recently Government of India has enacted a legislation to introduce crop insurance in the event of natural calamities like flood and drought. Flood risk zoning would help in a big way to provide crop insurance cover in such areas. In puts for preparing flood risk zoning are (i) flood inundation maps with respect to floods of different magnitude, frequencies / return periods in the historic past, (ii) land use in the flood plain (iii) existing and planned development schemes (iv) population (v) economic activities (vi) legislation (vii) building by laws and (viii) insurance policies. An attempt of NRSA flood risk zoning of Kosi river in Bihar based on limited satellite data is cited as an example in the **Figure** 7.



Once in 2 years Once in 60 years
Once in 150 years Drainage congested area

Figure 7 Flood risk zone map of Kosi river (Source : NRSA)

4.1.2 Capturing the spatial variability of cropping pattern

Spatial variability of cropping pattern get reflected in terms of spectral variations observed on the satellite imageries, if spatial, spectral, radiometric and temporal resolutions are adequate. Capturing these variability enables monitoring the state and conditions of the agricultural ecosystem. Cropping pattern of an area reflects the spatial arrangement of crops and intercrop variability that exists. Multi-date satellite data is essential to derive cropping pattern as in general there are more than one crop grown in a season with different growing calendar. Satellite remote sensing has number of advantages for monitoring cropping pattern such as (i) Multi-spectral observations which allows crop discrimination and growth assessment, (ii) Synoptic coverage at multiple spatial resolutions which leads to micro and macro level assessments over large areas (iii) Multi-temporal, *i.e.* multiple observations in a crop season (iv) Multiple utility *i.e.* same data can be used for many applications and (v) Full data archival and reanalysis capability *i.e.* historical primary data can be reanalyzed with new models as well as for post-harvest validation.

Cropping pattern analysis of Gangetic active flood plain of Bhagalpur district, Bihar extending from 25^0 12′ 50″to 25^0 24′ 31″N latitude and 86°43′37″ to $87^0$13′57″ E longitude was done using WiFS data of IRS-1D of the four dates (November 30, 1999, January 19, 2000, March 09, 2000 and April 03, 2000) during post flooding period (i.e. rabi season). The spatial resolution of WiFS sensor is 180m and it has two spectral bands *i.e.* red (0.77 to 0.86μm) and NIR (1.55 to 1.77μm). Along with multitemporal image registration, radiometric correction was carried based on pseudo-invariant features in the images.

The Normalized Difference Vegetation Index (NDVI) being a potential indicator for crop growth and vigour was used in study. The mathematical expression for NDVI is given as

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})}$$

Based temporal NDVI profile derived from images of different dates different crops and their spatial distribution was computed using ground truth data collected using GPS (Fig 8). It was found that out of about 2.27 lakh ha of study area, various agricultural crops cover 59 per cent. The wheat crop was found to be dominating one (20%) followed by maize (12 %) and lentil and chick pea (10 %). In wheat growing areas, about 2, 12 and 6 percent area was occupied by early, timely and late sown wheat respectively. Similarly, in maize crop, 8 and 4 percent was in early and late sown categories respectively. Winter maize and wheat is grown majorly in flood plains close to river bank (locally known as diara land), pulses are common in low lying areas called *tal* lands. Arhar crop sown during May/June if not damaged by flood remains in the field upto late March /early April. Other 41 per cent includes built up area, orchards/plantation areas, barren lands, river, water bodies etc. Digitized map of Bhagalpur with village boundaries was also overlaid to retrieve the information about late planted crops and fallow lands for possible interventions.

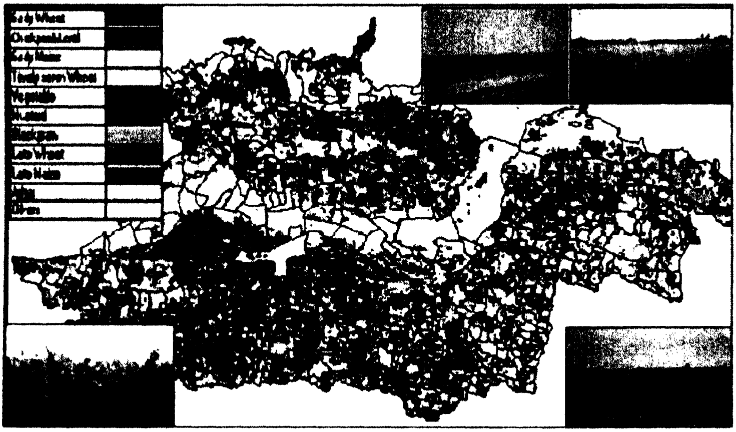


Figure 8 Spatial distribution of crops grown (*rabi,*1999-2000) overlaid with village boundaries of Bhagalpur district, Bihar

4.1.3 Addressing the seasonal dynamism of agricultural system

Dynamism of crops over an agricultural calendar can be captured from 10 days maximum value composite satellite data products like NDVI images of SPOT-4 vegetation sensor. NDVI data from a satellite sensor is primarily related to vegetation changes and follows annual cycle of growth and decline. Thirty six NDVI images of 10 days composite for the agricultural calendar year June 1999-May 2000 were stacked together to derive temporal NDVI profile. Clouds and poor atmospheric conditions usually depress NDVI values causing sudden drops in NDVI which are not compatible with the gradual process of vegetation growth and is considered as noise. This noise was removed using Fourier Adjusted Sun Zenith Angle Corrected Interpolated and Reconstructed (FASIR) Technique. Corrected image was used to retrieve smooth temporal NDVI profile, were analyzed using ground information and whole area was classified into four classes of depending on cropped and fallow lands during *kharif* and *rabi* seasons (Fig .9).

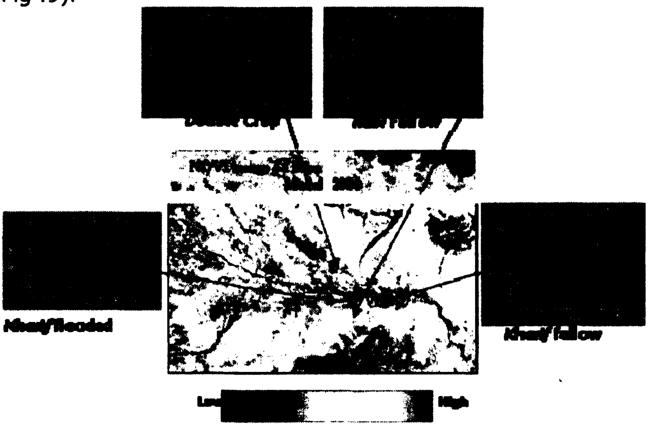


Figure 9. Stacked SPOT NDVI image showing temporal NDVI profile of a year for kharif fallow, rabi fallow, kharif flooded and double cropped area.

4.1.4 Targeting for possible interventions to enhance the productivity

Satellite remote sensing along with Geographic Information System helps in retrieving and integrating various information to make possible decision on intervention for enhancing productivity at spatio-temporal scale. For example spatio-temporal pattern of flood and crops in flood plain agro-ecosystem indirectly gives insight for targeting technologies. As discussed in the previous sections, remote sensing could be used to know potential flooding areas, its recession pattern and land availability for cultivation which decides the kind of cropping pattern practiced in the region.

Many lands having high production potential remains fallow due to late harvesting of *Kharif* crop and longer turn around time between two crops in the system. In upland flood plain areas where risk of flood is minimal and land is available through out the year, coarse or fine rice crop is taken during *Kharif* season. Timely harvest of coarse rice leads to timely sowing of wheat crop. But the late harvest of fine rice during December results in late planting of wheat crop, which is not profitable for farmer. Hence, they either prefer to keep the land fallow or grow vegetable crops in some patches. Land remaining fallow after harvest of rice in kharif is called rice-fallow. Inventory on rice fallow made by NRSA and ICRISAT through IRS WiFS data analysis during 1999-2000 revealed that India alone has 11.65 m.ha, which is 29 % of rice area. Nearly 82 % rice fallows are located in the most flood prone states. The GIS analysis of these fallow lands has indicated that they represent diverse soil types and climatic conditions, thus crop diversification through growing varieties of warm and cool season legumes can be best suitable option having multi-faceted impact on the economy. In such areas also mechanization of crop harvesting and use of different Resource Conserving Technologies (RCTs) for timely planting would bring the fallow lands under cultivation and can significantly increase the productivity of the system. Resource Conserving Technologies include zero, minimum and reduced tillages, surface seeding etc which not only save turn around time, reducing land preparation time but also increase water and nutrient use efficiencies

Again, generally flooding period in Eastern Gangetic plains (namely Bhagalpur district of Bihar) ranges from 3rd week of July to 4th week of September. In these areas sowing schedules of crops (mainly wheat and winter maize) depends exclusively on the recession pattern of flood, which decides the availability of land for cultivation. Early recession of flood allows farmers to take black gram followed by either late wheat or winter maize and summer maize. Black gram is a bonus crop as it is broadcasted in the fields to grow on residual soil moisture, there by it does not require any land preparation and inputs except the seeds. But this crop is harvested in December which delays planting of succeeding wheat and winter maize, consequently affecting its yield and also indirectly secure harvest of the summer maize before onset of next flood. Traditionally, in these areas, wheat crop is sown by broadcasting method after five ploughings at the seed rate of 250 kg ha^{-1} and basal fertilizer is broadcasted on the surface. Due to ploughing operations, late planting of wheat crop results in poor yields due to reduction in growing period of the crop. To avoid late planting and reduce the seed rate and for proper placement of basal fertilizers in the root zone, seed cum fertilizer drill (*i.e* one of the RCTs) could be a potential tool. From the exploratory trial in farmers fields in diara lands of Sabour block, Bhagalpur under aegis of Rice Wheat Consortium, CIMMYT, New Delhi, it was found that seed-cum-fertilizer drill could reduce commonly practiced five ploughings to two which were good enough for proper placement of seed and fertilizers and reduced the seed rate by about 150 kg ha^{-1}. This also resulted in producing long and healthy ears and hence the higher yields compared to traditional system.

4.2 Forecasting and management of drought

Remote sensing techniques have potential application in three major aspects of agricultural drought: (a) early warning (b) monitoring and (c) drought proofing and management. Early warning of impending drought is intimately linked with long and medium drought is intimately linked with long and medium range weather forecasting particularly that of monsoon, failure of which is the cause of drought in India.

Under the National Agricultural Drought Assessment and Monitoring System (NADAMS), drought assessment with district as a reference unit, and monitoring through the Kharif and part of rabi seasons (June-December), is primarily based on normalized difference vegetation index (NDVI) derived from NOAA satellite. Advanced Very High Resolution Radiometer (AVHRR) sensor data are acquired daily at the earth surface. The NDVI is a function of green leaf area and biomes and is indicative of the vegetation status. The drought severity classification-mild, moderate and severe is based on criteria in regard to vegetation index anomalies compared to the previous normal year. The persistence of drought conditions is indicated by the progressive drought status table.

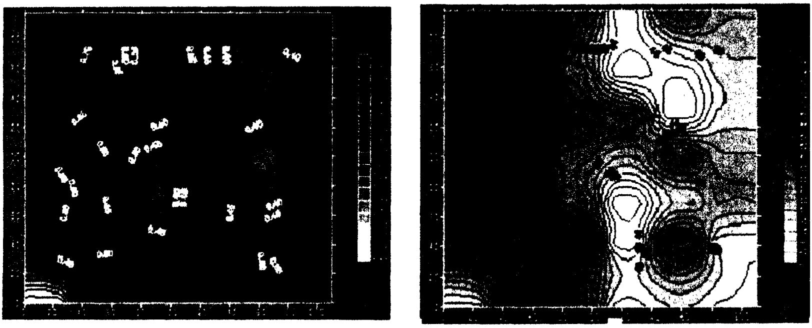
Remote sensing techniques in conjunction with geographic information system (GIS) hold great promise for drought proofing or management on short term as well as long term basis. By incorporating socio-economic data coupled with information on soils and crops, packages are developed for integrate management of agricultural land, water and fodder resources. Maps in 1:50,000 scale having themes, such as land use, soils, wastelands, surface waster bodies, ground water prospect sites, forest type and crops etc. are prepared and used for integrated drought management programme.

5 Remote Sensing for Precision Farming

Advancement of Remote Sensing with other spatial technologies in 20th century led to the development of the concept of precision farming which relies on the integration of these technologies into a single system that can be operated at farm level with sustainable effort. Precision Farming (PF) which is otherwise known as Site-Specific farming involves matching resource application and agronomic practices with soil properties and crop requirements as they vary across a site. PF has three requirements such as (i) ability to identify each field location, (ii) ability to capture, interpret and analyze agronomic data at an appropriate scale and frequency, and (iii) ability to adjust input use and farming practices to maximize benefits from each field location. Collectively, these actions are referred to as the "differential" treatment of field variation as opposed to the "uniform" treatment underlying traditional management systems. The result is an improvement in the efficiency and environmental impact of crop production systems.

Remote sensing holds great promise for precision agriculture because of its potential for monitoring spatial variability over time at high resolution (Moran *et.al.,* 1997). Various workers (Hanson *et al.,* 1995) have shown the advantages of using remote sensing technology to obtain spatially and temporally variable information for precision farming. Keeping in view the agricultural scenario in developing countries, the requirement for a marketable RS technology for precision agriculture is the delivery of information with the following characteristics like low turn around time (24-48 hrs), low data Cost (~ 100 Rs./acre/season), high Spatial Resolution (at least 2m multi-spectral), high Spectral Resolution (<25 nm), high temporal Resolution (at least 5-6 data per season) and delivery of analytical products in simpler format.

Variable rate technology in PF is key component which is done for site specific input management and it is done in two ways The first method, Map-based, includes the following steps: grid sampling a field, performing laboratory analyses of the soil samples, generating a site-specific map of the properties through geostatistical technique (**Figure** 10) and finally using this map to control a variable-rate applicator. During the sampling and application steps, a positioning system, usually DGPS (Differential Global Positioning System) is used to identify the current location in the field. The second method, Sensor-based, utilizes real-time sensors and feedback control to measure the desired properties on-the-go, usually soil properties or crop characteristics, and immediately use this signal to control the variable-rate applicator



(a) (b)

Figure 12. Spatial variability map of (a) Soil organic carbon and (b) grain yield in (q ha^{-1}) of wheat crop in 4.16ha land in IARI farm.

Precision Farming thus calls for the use of appropriate tools and techniques, within a set of the framework as mentioned, to address the micro-level variations between crop requirements and applications of agricultural inputs. Inevitably, it integrates a significant amount of data from different sources; information and knowledge about the crops, soils, ecology and economy but higher levels of control require a more sophisticated systems approach. It is not simply the ability to apply treatments that are varied at the local level but the ability to precisely monitor and assess the agricultural systems at a local and farm level. This is essentially to have sufficient understanding of the processes involved to be able to apply the inputs in such a way as to be able to achieve a particular goal not necessarily maximum yield but to maximize financial advantage while operating within environmental constraints.

6. Remote Sensing for Farmers Advisory Services

Proper dissemination of timely and relevant information is essential for the farmers. Such information would encompass not only the structured spatial information derived from the satellite remote sensing but also information related to package of practices, agricultural inputs (seeds, fertilizers & plant protection chemicals) and implements, subsidy, credit & insurance schemes, marketing infrastructure, agroclimatic trends, soil testing facilities etc. In this context , 'Agroclimatic Planning and Information Bank (APIB) generated for selected districts of Karnataka envisages to provide single window access for information related to agriculture and allied sectors through such detailed database. APIB is aimed at providing value added information to the decision makers at all levels. Initially taken up at the behest of the Planning Commission's Agro-Climatic Regional Planning (ACRP) project, and later for the Government of Karnataka, APIB has also been extended to a few districts in West Bengal and Meghalaya.

7. Village Resource Centres

Village Resource Centres (VRC), meant for effectively delivering a portfolio of services emanating from the space systems, directly down the line to the rural people. VRC is in fact the product of the learning and experiences gained through the various INSAT and IRS based projects carried out in the past. Technologically, VRC is based on the concept that the satellite applications, both communication and remote sensing could be dovetailed and multi-tasked to disseminate a variety of relevant community-centric services, including tele-medicine and tele-education. The VRCs are envisaged to play an enabling role in addressing the primary issues related to eradication of illiteracy, better healthcare, training on vocational jobs, providing agricultural advisories, and helping to enhance agricultural productivity.

The VRCs, is essentially a step towards evolving innovative, cost effective and community-based local contextual solutions integrating the space-enabled services seamlessly with other locally relevant services. Eventually, with the increased involvement of local community, the VRCs would also be able to facilitate e-Governance at local levels; and will also promote rural entrepreneurship. VRCs are essentially set up in partnership with NGOs, Trusts, as well as concerned State and Central agencies.

8. Input/Output management/ Post harvest practices

Timely information is also needed for optimum movement of fertilizers, pesticides as well as the seeds themselves, depending on the agricultural practices and the weather conditions. Space technology inputs along with Geographic information System (GIS) can be effectively used in input optimization and output management especially for post harvest practices in agriculture. In this direction, a decision support system for assessing the potential fertilizer requirement and their optimum movement has been developed for Indian Farmers Fertiliser Co-operative Ltd. (IFFCO). Space technology derived inputs have been effectively used in assessing the taluk-wise fertilizer requirement, while weather data have been used for assessing the time of requirement. Such integrated database was used in optimizing the route for movement of fertilizer from the source to the sale points using the digital road network and location of IFFCO's storage infrastructure. Such studies can also be replicated in reducing post harvest loss, which is significant in vegetables and fruits.

9.Conclusions

Agriculture systems to survive,have to be ecologically sound and more profitable by virtue of enhanced productivity at lower cost of cultivation. It needs to be therefore transition to a more information and knowledge based so that it could address not only the issue of poverty alleviation but also serve as an engine to the accelerated growth in the region. Geoinformatics, the integrated technology of Remote sensing, GIS and GPS is found to be most suited to make agriculture to be information and knowledge and to monitor also associated yield reducing environmental events so that timely generated data and information extracted can further be used for strategic improvement of agricultural productivity in a stress environment.

References

Burrough, PA.(1986), Principles of Geographic Information Systems for land Resources Assessment, Clarendon Press, Oxford

Goel N. (1989), Inversion of canopy reflectance models for estimation of biophysical parameters from reflectance data, In *Theory and Applications of Optical Remote Sensing* (G. Asrar, Ed), Wiley Interscience, New York, pp. 205-250.

Hoffmann-Wellenhof, B. H. Lichtenegger, and J. Collins. 1998. *GPS: Theory and Practice*. 4th ed.New York: Springer-Verlag, 389p.

Jacquemoud, S and Baret, F (1990) 'PROSPECT: A Model of Leaf Optical Properties Spectra, *Remote Sensing of Environment*, 34: 251-256.

Jensen, J.R. 1996. Image Preprocessing: Radiometric and Geometric Correction. In : Introductory Digital Image Processing. Prentice Hall publisher (Second Edition): 107-135.

Kennedy, A. 2002 *The Global Positioning System and GIS : An Introduction* 2nd ed. Taylor and Francis, 345p.

Leick, Alfred. 1995. *GPS Satellite Surveying*. 2nd. ed. New York: John Wiley & Sons.

Moran, M.S., Inoue, Y. and Barnes, E.M. (1997)Opportunities and limitations for image based remote sensing in precision crop management. Remote Sensing Environment, 61:319-346.

Nagarajan, S., Dadhwal,V.K. and Sahoo R.N. 2004. Role of earth observation technology and applications in India's food security Concerns, *Proc. on India-United States Conference on Space Science, Application and Commerce - Strengthening and Expanding Co-operation*, held at Bangalore, June 21-25, 2004 : 25.

Parkinson, Bradford W. and James J. Spilker. eds. 1996. *Global Positioning System: Theory and Practice. Volumes I and II.* Washington, DC: American Institute of Aeronautics and Astronautics, Inc.

Sahoo, R.N. and Arora, R.P. 2001, Precision farming in Indian agriculture. Proc. Natl. Symp. on *'Recent Advances in Remote Sensing and GIS Technologies for Natural Resources Management'* IIT, Mumbai, December 5-7, 2001: 25.

Sahoo, R.N., Tomar, R.K. and Arora, R.P. 2002. Precision farming: A prospective in 21st century. Proc. 2nd Int. Agronomy Congress on *Balancing food and environmental security - A continuing challenge*, IARI, New Delhi, Nov. 26-30, 2002:1194-95.

Sahoo, R.N., Tomar, R.K., Sehgal, V.K., Rao, C.S., Nirupa Charchi., Abrol, I.P., Wadhwani, M.K. and Gupta, R.K.(2004). Geo-informatics for enhancing crop productivity in Eastern flood plains-A case study. Proc. Natl. Symp. on *"Geo-informatics Applications for Sustainable Development"* Feb. 17-19, IARI, New Delhi : 183-189.

Sahoo, R.N., Tomar, R.K., Abrol, I.P., Krishnamurty, V.V.R.S., Wadhwani, M.K.,and Tiwari, M.K.(2002): Drivers of land use dynamics in Gangetic plains: Application of Remote sensing and GIS analysis. Proc. LUCC-ISPRS workshop *"LUCC Contribution to Asian Environmental Problems*, Dec. 2, NRSA, Hyderabad, pp. 20-21.

Sahoo, R.N., Tomar, R.K., Rao, C.S., Abrol, I.P., Wadhawani, M.K. (2003). Potential of RCTs in Eastern Gangetic Flood Plain: A Remote Sensing Analysis, In Rice-Wheat Information Sheet, Rice Wheat Consortium for the Indo-Gangetic Plains (RWC), CIMMYT, New Delhi, (Rice Wheat Information Sheet, RWIS), 47:8-9.

Sahoo, R.N., Tomar, R.K., Rao, C.S., Sehgal, V.K., Charchi, Nirupa, Abrol, I.P., Tiwari, M.K. and Wadhawani, M.K. (2003). Radiometric Scene Correction of Temporal Multispectral Satellite Data for Discrimination of Crops, *Indian Journal of Radio and Space Physics*. (in Press)

Venkatachary, K.V, Bandyopadhaya, K., Bhanumurty, V, Rao, G.S., Sudhakar, S. Pal, D.K. Das, R.K., Sharma, Utpal, Manikam, B, Meena Rani, H.C. and Srivastava, S.K. 2001 Defining a space based disaster management system for floods: A case study for damage assessement due to 1998 Brahmaputra Floods, Current Science Vol. 80 (30) pp. 369-377.

TECHNOLOGICAL INFORMATION ON ORNAMENTAL FISHES- A RESOURCE FOR ENTREPRENEURS

S. K. Swain[1] and N.K. Barik

Abstract

Information is the resources as the entrepreneurs with right technological information can earn more by setting ornamental fish unit in the rural as well as urban centers. Fish keeping today is the world's most popular hobby after photography and ornamental fish are the most popular pets in the world. Keeping ornamental fish and its propagation provides not only aesthetic pleasure but also an excellent employment opportunity to many especially for unemployed youth, women and others through construction of small-backyard units. It has got a tremendous export potential for earning foreign exchange for the country. In this article attempt has been made to discuss about various aspects culture and some model economics of breeding and culture of ornamental fishes. The status of ornamental fish breeding and culture in the country is reviewed.

1. Introduction

Ornamental fish farming is a promising sector within aquaculture which creates ample opportunities in terms of the growth, generating income and employment to the large number of the unemployed youth in the country side. At present only a fraction of the domestic and international potential is harnessed. But, in recent times, the sector has shown faster growth due to concentrated efforts of the farmers and entrepreneurs to take up the ornamental fish farming as mean of their business and livelihood. A large number of the stakeholders' i.e. fishers, farmers, breeders, traders, vendors, transporters and exporters are involved in the sector. The whole business of the ornamental fish trade is based on the supply of the fishes from two primary sources *i.e.* wild collection and captive breeding. There is a wide apprehension on the environmental impact of the wild collection and it has a damaging effect on the threatened fish biodiversity in the country. Therefore, captive breeding is the primary approach for the sustainable development of the enterprises. The breeding not only reduce the pressure from the nature to a considerable degree, but provides in-vivo means of germplasm conservation through culture. Therefore, development of the breeding technology is indentified as the critical and priority for sustainable development of the sector. The paper reviews the development of the breeding technology of the ornamental fishes in India with some salient aspects of the technology.

ORNAMENTAL FISH SPECIES FOR COMMERCIAL PRODUCTION

Important groups of egg-layers are barbs, rasboras, goldfish, tetras, danios, bettas and gouramis and the major livebearers are guppies, platies, mollies and swordtails. Barbs are the most important group and most species of the group are known to have originated from India, *viz.*, rosy barb (*Puntius conchonius*),

[1] **Correspondence:** Saroj Kumar Swain, Principal Scientist, CIFA, Bhubaneswar-751002, Email: sarojswain2008@gmail.com or swain_saroj2002@rediffmail.com

melon barb (*P.fasciatus fasciatus),* aruli barb (*P. arulius*). *etc*. Among the exotic barbs, tiger barb (*P. tetrazona*) and its aquarium developed strains like green tiger barb and red tiger barbs are fascinating in modern days. The major species of the group danios include zebra danio (*Brachidanio rerio*), pearl danio (*B. albalineatus),* giant danio (*Danio aequipinnatus*), turquoise danio (*D. devario*), and malabar danio (*D. malabaricus*). Zebra danio is one of the typical examples of ornamental fish species of Indian origin, which can be bred and reared easily. Among rasboras, the slender rasbora (*Rasbora daniconius*), glowlight rasbora (*R. pauciperforata*), harlequin rasbora (*R. heteromorpha*) and scissors tail (*R. trilineata*) are important ones. The gold fish, *Carassius auratus*, is the most commonly available fish, preferred by the hobbyists because of its attractive colouration ranging from pure gold to red, orange, black and albino. Some of the common varieties of gold fish available are comet, lion head, oranda (a modification of lion head), fringe tail, veil tail, fantail, shubunkin, telescopic eye, etc. These fishes grow up to 20 cm in length, but start breeding when they are only 6 cm long. Many aquarium breeds strains of ornamental koi, *Cyprinus carpio* (var. koi) easily by available hormone injections. Its origin is from East Asia but strains like orenji ogon, kin matsuba, kohaku, and shiro-bekoo are being cultured worldwide.

The tetras are small fishes of 3-8 cm long; majority of which has originated from South America. The most common species of the group are black widow tetra (*Gymnocorymbus ternetzi*), serpae tetra (*Hyphessobrycon sarpae*), lemon tetra (*H. pulchripinnis*), flame tetra (*H. flammeus*), neon tetra (*Paracheriodon innesi*), cardinal tetra (*P. axelrodi*), black neon tetra (*Hyphessobrycon herbertaxelrodi*) and pretty tetra (*Hemigrammus pulcher*). The species, *Betta splendens* commonly called as Siamese fighting fish occurs in varied colours like green, red, blue, albino and sometimes with a combination of two or three shades. The attractive colour and hardiness of the species are the characters of the species for its wide adoption by the hobbyists. The males are brightly coloured with beautifully spread-over fins. They show aggressive behaviour, only when other males are present. Angel fish, Pterophyllum *altum* and *P. scalare* are important candidate species widely preferred for aquaria, with different developed varieties such as black, veil tail, gold and black marble, platinum, gold pearl, pink ghost and albino. The filamentous lower fins and their compressed body shape with their elegant movements are pleasant to look at. Among gouramies, three spot gourami (*Trichogaster trichopterus*), pearl gourami, (*T. leen*), moon light gourami (*T. microlepis*), snakeskin gourami (*T. pectorails*), dwarf gourami (*Colisa lalia*), honey gourami (*C. sota*), striped gourami (*C. fasciata),* noble gourami (*Ctenops nobilis*) and kissing gourami (*Holostoma temmincki*) are the important species. The bettas and gouramies are most popular among nest-builders, characterized by possession of accessory respiratory organs, thus making the species hardier. Among catfishes the exotic variety of sucker mouth catfish (*Hypostomus multiradiatus*) is very popular, being originated from South and Central America it is available in every aquarist.

2. Salient features of production of ornamental fish

The success of any entrepreneurs depends upon the project planning, site selection and successful layout, design of the breeding or rearing unit. Once the unit is established in any site and latter on found uneconomical due to unavailability of certain important facilities like water, power etc. can not be rectified in latter stage. Therefore, proper planning is required before entering in to such flourishing farming practice. At present the variety of commercial enterprises producing ornamental fishes is as wide as the species produced. The degree of intensification and species farmed depends on following aspects.

- Training on the subject is a prerequisite before starting an ornamental fish unit.
- The minimum land requirement is 500-1000 square feet area for a small scale farming practice, whereas 1 acre and more for large scale farming in which few earthen ponds are to be excavated for some species like koi carps, gourami, barbs etc.
- Site selection is one of the main criteria where the farmer should select a cool environment for the culture and breeding.
- Breeding and rearing unit should be made near a constant supply of water and electricity.
- In cold climates this farming is too expensive to make the water warm to produce tropical fish. Where as the tropical climates favour the production rate because of year round breeding, rearing and better growth. So the entrepreneur has to select the fish accordingly.
- The selection of candidate species depends on the water quality of that area.

Because there are species prefers either soft water or hard water for breeding. There are species they can withstand a wide range of water condition tolerating all kinds of water. For example all the live bearers prefer hard water- alkaline. Egg layers like Goldfish, Gourami, Danio, Catfish, Rosy barb, Fighter etc. can tolerate wide range of water condition. Species like Angel, Discus, Tetras, Oscar, Loach etc. prefers soft -acidic water. Therefore, water quality can be checked in any nearby water chemistry laboratory.

- In certain areas, some fishes are difficult to breed, where there is a severe problem of water quality in any extreme conditions of physico- chemical conditions. Therefore, we should avoid such areas for establishing ornamental fish farms.
- Biofiltration unit is a prerequisite for smooth functioning of an ornamental fish culture and breeding unit.
- The brood stock selected for breeding should be of superior quality, so that good quality fish seed could be produced. If brood stocks are not available in that area, one can think of rearing a desired quantity of smaller fish and further they can be developed in to brood stock.
- Brood stocks can be allowed to breed for not more than two years. Fresh stocks from different source may be added in every two years to the selected parent stocks to improve the breeding efficiency and produce healthy offsprings.
- The fish breeder should concentrate preferably on one species so that it helps the breeder to develop expertise on the particular species and a good variety of fishes can be produced as per the market demand.
- Constant availability of agro-based byproducts will facilitate preparation of pelleted diet for the fish. For preparing a pelleted diet a mini pelletiser can be installed.

- The breeding and rearing unit may be established preferably nearer to air port/railway station, bus stand etc. for easy transportation for export and domestic market
- The breeders should develop market relations with pet/ retail shops, potential farmers, vendors dealing with ornamental fish, marketing network, etc. to facilitate the process of selling/procuring new brood stocks.
- A committed entrepreneur should always ensure regular contact with the recent research developments in the field and attend training and exposure visits.
- All new incoming fishes should be quarantined from resident stock. Movement of fishes should be restricted from a suspected or unknown disease status area.
- Few quarantine tanks are required little away from the unit so that a proper observation can be made on heath aspects.
- If any abnormal behaviour is observed in any culture tanks then the fishes needs to be isolated immediately. If mass mortality occurs in the farm then experts needs to be consulted before application of any treatment.

3. Important aspect of water quality

The degrees of hardness have several biological effects upon aquatic life. Bicarbonates tend to prevent a solution from changing acidity. Soft water, lacking this protection, may become particularly acidic when much carbon dioxide is present; such a change creates stress for organisms. For soft water species excessive hardness causes an organism problem in absorbing substances through its delicate membranes. This is most true of the sensitive naked cells of eggs and milt, so that soft water has been found to play a vital role in the successful reproduction of many species of freshwater fishes. Thus, at least for purposes such as fish breeding, a soft solution is desirable.
To maintain soft water, all sources of calcium carbonate such as calcareous rocks, gravels, coral, broken shell and algae must be kept out of the aquarium system whilst using only soft water initially and during exchange. Conversely, presence of such sources will preserve the water hardness. Some of the important water quality parameters and their optimum ranges for tropical aquarium fish are presented in the table below:

Table-1 water quality parameters for ornamental fish

Temperature	24-28°C
pH	7.0-8.5
CO_2	< 5 ppm
Alkalinity	75-120 ppm as $CaCO_3$
Hardness	60-100 ppm as $CaCO_3$
Dissolved oxygen	6.0-8.0 ppm
Free ammonia	<0.05 ppm
Ionized ammonia	<0.1-0.4 ppm

4. Production system

There are two types of production system existing in any aqua-farming practice. First type is production through breeding and larval rearing up to 2-3 months and that could be possible by constructing cement tanks of various sizes meant for breeding and larval rearing. The second type is culture/ farming to produce fish from larvae to marketable size for a period of 3-6 months of rearing either in cement tanks or in pond culture system depending upon the species. The farmer may go for any one of the above production system for better result. Accordingly, the farmer or entrepreneur has to decide to take up the breeding/culture or both as per his skilled technical manpower. However, the small scale farmer in a back yard dealing with small fishes like livebearers may not come in the above categories.

4.1 Pond cultures of some ornamental fish- Some of the ornamental fish can be cultured in earthen ponds, the culture practice being followed are similar to conventional fish culture. The technique is extensive and is characterized by low stocking densities and with a heavy emphasis given on live food (plankton) production. Due to a low labour requirement, pond culture is one of the cheapest production methods but the entrepreneur enjoys limited control over the culture system. The culture system in such a case must be provided with proper guard against predating animals such as cormorants and snakes. Small water bodies such as tanks, cisterns and small sized lined pond are most suitable for 'pond' culture of ornamental fish. Small tanks or cisterns are suitable for breeding purpose. For rearing, larger sized tanks/cisterns are preferable. Small sized lined ponds are suitable for raising of brood-stock. Most significantly, the size of tank/cisterns/ponds depends on the different species to be cultured, though in certain cases breeding/culture practice being employed could also be considered. However, the species like Koi carps, Comet gold fishes, Gouramies, Cat fishes and other hardy fishes can be cultured in earthen tanks, provided all facilities for controlling the predators are existing.

5. Economics of ornamental fish breeding and culture

The economic viability is the foundation of the any popular and successful enterprises. On the basis of the business scale, the ornamental fish farms can be categorized as small scale or large scale. The large scales are further divided into live bearers and egg layers units. The small scale units are low investment backyard enterprises with low intensity of production where as the large scale commercial enterprises involve higher cost and returns. The detailed economics of three categories of the ornamental farms are presented below.

Table-2.Economics of small-scale breeding and rearing unit for live-bearer ornamental fish species

		Amount (Rs.)
	1. a) Fixed Capital	
1	Low-cost shed of 300 sq. ft. area (bamboo frame with net covering)	10,000
2	Breeding tank (Concrete) 6' x 3' x 1'6" @ Rs. 2500/tank (4 nos.)	10,000
3	Rearing tank (Concrete) 6' x 4' x 2'0" @ Rs. 2800/tank (2 nos.)	5,600
4	Brood stock tank (Concrete) 6' x 4' x 2'0" @ Rs. 2800/tank (2 nos.)	5,600
5	Larval tank (Concrete) 4' x 1'6" x1'0" @ Rs.1200/tank (8 nos.)	9,600
6	Bore-well facility with 1 HP Pump	8,000
7	Oxygen cylinder with accessories (1 no.)	5,000
	Sub-total	**53,800**
	b) Recurring Expenditure (for one year)	
1	Brood fish of guppy, molly, swordtail & platy (800 female & 200 male @ Rs. 2.50/pc.)	2,500
2	Feed 150 kg/year @ Rs. 20/kg	3,000
3	Different types of nets	1,500
4	Electricity and fuel @ Rs. 250/month	3,000
5	Wages for labour (@ Rs. 1000/month)	12,000
6	Miscellaneous (chemicals, medicine, packing material, gas cylinder/filling, perforated plastic breeding basket etc.)	2,500
	Sub-total	**24,500**
	2. Total Expenditure	
1	Recurring cost	24,500
2	Interest on fixed cost (12% per annum)	6,480
3	Interest on recurring cost (12% half yearly)	1,470
4	Depreciation (20% of fixed cost)	10,760
	Total	**43,210**
	3. Gross Income	
	Sell of 76,800 nos. fish @ Rs.1/pc, reared for one month (@ 40 nos./female/cycle from 3 cycles/year, and considering survival of 80%)	76,800
	4. Net Income (Gross income-Total expenditure) **Rs. 76,800-43,210**	**33,590**
	5. Monthly income	**2800**

It is estimated that, within a year's time of three breeding cycle, a beginner may earn a sum of Rs. 2800 per month. If a farmer does the breeding and rearing work single-handedly, he/she can add the labour cost into the income. Upon acquiring further expertise, the farmer can increase his/her income by many ways, such as by increasing the number of breeding cycle per year, standardizing the overall hatchery procedures that would further reduce the expenditure on feeding, and mastering upon the breeding techniques to increase fry/young ones production in each cycle. Therefore, the monthly income of the three thousand rupees is the most conservative estimation of the economics of the enterprises. The farmers can earn up to 10000 rupees per month with the same infrastructure by developing expertise, improving efficiency and better price. This income can be considered to be high compared to any other similar enterprises.

Table-3.Economics of large-scale ornamental fish breeding and rearing unit for live-bearers

		Amount (000 Rs)
	a. Fixed Capital	
1	Land required 1 ha	-------
2	Shed with electrical supply (4000 sq. ft.)	800.0
3	Breeding tanks (Concrete, 6' x 3' x 2', 24 nos @ Rs. 2500/tank)	60.0
4	Rearing tank (Concrete, 8'x4'x2' @ Rs. 4000/tank; 30 nos.)	120.0
5	Brood stock tank (Concrete 6'x3'x2' @Rs. 2500/tank; 40nos.)	100.0
6	Larval tank (Concrete 3' x 2 x1' @Rs.1000/tank; 20 nos.)	20.0
7	Pump, overhead tank and water supply facility	150.0
8	Air blower and other aeration systems	150.0
	Sub-total	**1400.0**
	b. Recurring expenditure	
1	Land lease	10.0
2	Brood fish (6000 female, 2000 male @ Rs. 5.00 /pc for each livebearer (potential reproductive period two years)	37.5
3	Feed (600 kg/year @ Rs. 40/kg)	25.0
4	Nets and other materials	25.0
5	Wages for labourer (@ Rs. 3000/month x3 nos)	108.0
6	Miscellaneous	25.0
	Sub-total	**230.5**
	2. Total Expenditure	
1	Recurring expenditure	230.5
2	Depreciation (10% of capital cost)	140.0
3	Interest on capital cost (12%)	168.0
4	Interest on recurring expenditure (12% for half yearly)	13.3
	Total	**551.8**
	3. Gross Income	
1	Fish sale (for 6,48,000 nos. @ Rs 1.50/piece) (Considering 90% survival from 7,20,000 seed produced from 4000 female in 3 times in a year)	**972.0**
	4. Net Income (Gross income-Total expenditure)	**420.2**
	5. Monthly income	**34.0**

Table-4. Economics of large-scale ornamental fish breeding and rearing unit for Egg layers

Species maintained: goldfish (Red cap, Oranda, Subhunkin and Black moor)

		Amount (000 Rs)
	a. Fixed Capital (Land requirement 1.5 ha)	
1	Shed with electrical supply (4000 sq. ft.)	800.0
2	Bloodstock tank (Concrete 6'x3'x2' @Rs. 2500/tank; 30 nos.)	75.0
3	Breeding tanks (Concrete, 5' x 3' x 2', @ Rs. 2000/tank; 30 nos)	60.0
4	Larval tank (Concrete 3' x 2 x1' @Rs.1000/tank; 50 nos.)	50.0
5	Rearing tank (Concrete, 8'x 4'x2' @ Rs. 4000/tank; 60 nos.)	240.0
6	Earthen ponds (0.05 ha 2 nos)/large concrete tanks (50 sq. m 6 nos)	300.0
7	Pump, overhead tank and water supply facility	200.0
8	Air blower and other aeration systems	200.0
9	Generator, gas cylinder and other requisites	150.0
	Sub-total	**2075.0**
	b. Recurring expenditure	
1	Land lease	15.0
2	Brood fish (3000 female, 4000 male @ Rs. 25.00 /pc)	175.0
3	Feed (1,500 kg/year @ Rs. 40/kg)	60.0
4	Nets and other materials	25.0
5	Wages for labourer (@ Rs.3000/month x 4 nos)	144.0
6	Miscellaneous including power, fuel etc.	100.0
	Sub-total	**519.0**
	2. Total expenditure	
1	Recurring expenditure	504.0
2	Depreciation (10% of capital cost)	207.5
3	Interest on capital cost (12%)	249.0
4	Interest on recurring expenditure (12% for half yearly)	31.0
	Total	**1011.0**
	3. Gross Income	
1	Fish sale (for 12, 00,000 nos. @ @ Rs 2/ piece) (Considering breeding success in 2000 female with average fecundity of 1000/fish once in a year 60% survival from 20,00,000 seed produced)	**2400.0**
	4. Net Income (Gross income-Total expenditure)	**1399.0**
	5. Monthly income	**115.5**

A successful economic enterprise requires lot of dedication, hard work, sincerity and timely marketing of the produce. So also in ornamental fish, the success depends on the investment, species selection, demand, and proper marketing. The model economic analysis of the various types of the ornamental fish farms presented here are indicative, and average estimation. The actual profitability varies from the place to place depending on the condition and efficiencies in the operation and management of the farm. As the enterprise involves live specimen, the involvement of the risk is higher. At the same time, the ornamental nature of the product provides opportunity to harvest comparatively higher prices than food fishes. The demand creation and arousal of the interest of the people can fetch the entrepreneurs' higher prices. Therefore it is always advisable to make a proper viable project prior to establishing any ornamental fish breeding unit with due consultation.

Considering the proven success of involvement of women in development of backyard enterprise in farming of ornamental fish in West Bengal and Kerala, it is necessary that due encouragement is given for creation of women SHGs for such enterprise. It may be interesting to note that, Malaysia with similar climatic condition like that of India has been able to grow the industry only with small backyard units at individual levels. Thus, development of ornament fish farming in any region, not necessarily requires very large-scale enterprises, but a few successful clusters of back yard units in different regions. In this regard special packages may be provided to these SHGs and unemployed youths or ex-service men for establishment of such enterprise.

6. New initiatives

6.1 Breeding indigenous species of ornamental value

Central Institute of Freshwater Aquaculture, Bhubaneswar has successfully bred 16 indigenous ornamental species from NEH, Eastern and Southern region. They are barbs, danios, rasboras, catfish, eel, black carp and chameleon fish. College of Fisheries, Panangad, Kochi has bred 15 numbers of indigenous ornamental fishes from Western Ghats. They are also barbs, danios and rasboras of Western ghat prioritized fish species. College of Fisheries, Raha, Assam has bred a few fishes from Assam, of which peacock eel is important.

6.2 Cities involved in trade and farming

Kolkata has been the largest exit point for ornamental fish export, followed by Mumbai and Chennai. These metropolitan cities are also the major breeding centres for freshwater ornamental fishes. In the recent years, people in the States of Kerala, Andhra Pradesh, Orissa, Bihar and other places have started producing ornamental fishes. According to an estimate, presently there are 350 full time and more than 2,000 part time fish breeders in the country. It has been observed that most of the aquarium breeding centers and the related trade have been confined in and around the metropolitan cities like Kolkata, Chennai and Mumbai, primarily due to ready urban market and availability of international airport facilities for import and export business.

6.3 Ornamental fish village at Udairampur in Amtala, West Bengal

This village is known to be an ornamental fish village, where over 200 families are involved in ornamental fish breeding as backyard activity with a market facility at a weekly market at Ghalib Street, near Hathibagan. There are about

500 commercial production units for ornamental fish in the districts of Howrah, Hoogly, South and North 24 Parganas in West Bengal.

6.4 Ornamental fish village at Kolathur, Chennai-
In Kolathur area of Chennai, people are involved in breeding and trade of different ornamental fishes and aquarium accessories. About 300 small to large scales units have come up in Kolathur area. Besides, there have been several other commercial farms established in different parts of Tamil Nadu, viz., Madurai and Tuticorin with emphasis on high valued species.

6.5 Urban area domestic units at Mumbai-
In Mumbai, there are some sophisticated breeders who breed rare and expensive fishes like Angels and Discus. There are more than 100 commercial units in Mumbai city and suburbs. It is interesting that the city has 5,000-6,000 small and part time breeders.

6.6 Ornamental fish units at Orissa-
Over 30 units have taken to ornamental fish culture as a backyard activity, with an investment of Rs. 10,000-Rs. 80,000/-. As many as eight varieties of ornamental fishes are bred by the units managed by individual families, with a monthly income of Rs. 2,000-5,000/-.

7. Opportunites

7.1 Possibility of development of such villages in different parts of the country
Ornamental fish farming and aquarium keeping was a hobby of the people in the high income group, are spreading to other strata of the society, because of the ease and affordability, as also innovations in the way the tanks are designed and fabricated. Backyard ornamental fish culture can be developed in and around cities like Bangalore, Cochin, Cuttack, Bhubaneswar, Tuticorin, Goa, Visakhapatnam, Hyderabad, Ahmedabad, Mangalore etc. Students and unemployed youth can very well be involved in ornamental fish breeding and culture, which will give them not only a regular monthly income but also means for self- reliance.

The Government of India provides financial assistance to the ornamental fish hatcheries, with a 10% subsidy component of Rs. 15 lakh per unit with a capacity of 5-10 million fry every year. The Marine Products Export Development Authority (MPEDA) and the National Fisheries Development Board (NFDB) also have provisions for supporting enterprises in ornamental fish culture and trade. There are provisions for the refinance facilities from the National Banks for Agriculture and Rural Development (NABRAD) which can be availed through any rural bank or commercial banks.

7.2 Role of Fisheries Research Institutes and Agricultural Universities
In the field of training and education, the ICAR Research Institutes such as Central Institute of Freshwater Aquaculture (CIFA), Bhubaneswar; Central Marine Fisheries Research Institute (CMFRI), Kochi; and Central Institute of Fisheries Education (CIFE), Mumbai; and Kerala Agricultural University are working on different aspects of freshwater and marine ornamental fish. CIFA is playing major role for conducting research and training on freshwater ornamental fish breeding

and culture in the country by conducting several national level training programmes both on and off-campus comprising more than 1,000 participants from all over the country during last 8 years. CIFA has developed a feed for some of the commercial varieties of ornamental fish containing 40% protein. Further, a pigment-enriched fish feed for rosy barb and gold fish for enhancing colours in fish has also been formulated. A medicine called CIFACURE for curing certain bacterial and fungal diseases has been commercialized and other medicines and vaccines are under pipeline. This Institute has also made a documentary film on 'Ornamental fish breeding and culture' in collaboration with SAARC Agricultural Information Centre (SAIC), Dhaka, Bangladesh for promotion and popularization of ornamental fish farming among SAARC member countries. The CIFA has also given the technical guidance for making two regional films on ornamental fish farming by UNDP and GRAMSAT, Bhubaneswar. Similarly for the North-eastern region, the Institutes regularly conduct the training programmes as also awareness programmes to conserve and breed the important ornamental species besides hatchery establishment in Meghalaya, Arunachal Pradesh, Nagaland and Assam.

8 CONCLUSION

There is great scope for developing small scale units with an investment of about Rs. 2 lakhs per unit towards both capital and operating costs. With backyard units comprising a few concrete tanks with a constant water source and seed, ornamental fish can be produced by farmer, unemployed youth and farm women. In a limited area of 500-1,000 sq. feet, they can earn Rs 2,000 to Rs. 5,000 per month with an investment of about Rs. 1 lakh. On a commercial scale, entrepreneurs have invested up to Rs. 10 lakhs, with a monthly net income of Rs 10,000 to 30,000. The practice is often a family enterprise, with the members joining hands in different activities of breeding, tank fabrication and maintenance, feed preparation, transport and sale, etc. From the existing level of 2,000 breeders, 1,000 retail outlets and 2,000 commercial units; in the country the present ornamental fish trade is around Rs 100 crores. This can be enhanced by three folds. Similarly, the export could be tripled with attention on the quality seed, pigment enhancement, feed development and health aspects. The Indian Council of Agricultural Research (ICAR) has taken steps for the production of quality ornamental fishes, both freshwater and marine, through its National Seed Project being operated in different centres of the country in last two years.

DATA GENERATION RESOURCES FOR BIOINFORMATICS

P. Das[1]

1. Introduction

The biological information that is being generated by research in modern biology worldwide is so huge that it is not possible to manage through conventional file keeping or simple computing. Secondly, analysis of these data certainly needs specific interfaces, software and management facilities. Therefore, requirement to store, analyze and manage the biological information/data in an appropriate way has led to the origin of a new branch of science called "Bioinformatics". Here, emphasis will be given on the research disciplines that mostly generates such data. Broadly, they can be classified as follows.

2. Genomics

Genomics deals with structural and functional analysis of genes in the genome. Often it is described under two sub-disciplines namely structural and functional genomics. Human genome was being expected to contain about 20 000 - 30 000 genes. The 15 years human genome project that got completed recently has elucidated that there are about 25000 genes. The objective of this project was to map and sequence all the genes in 23 pairs of chromosomes. The goals of the project have been divided into smaller tasks, one of which is to develop genetic linkage maps to establish markers on each of the 46 human chromosomes in order to work on smaller and more clearly defined sectors in the search for genes. A second task is to get a physical map of overlapping clones of large DNA fragments that will span all the chromosomes. The map with the highest resolution will show the location of genes or markers on the chromosomes as measured by the number of DNA bases between them. The ultimate goal is to get the entire DNA sequence of the human genome. The next challenge is to identify the functions of these genes.

Now we can have access to individual genes with its structural details much faster than ever before. Soon we will have an accurate function of all the genes. Likewise, the genome sequences of more than fifty prokaryotic and eukaryotic organisms have been completed. As said earlier, the primary goal of a mapping program is to construct genetic linkage maps sufficient to identify either a disease gene or chromosomal segments that harbor quantitative trait loci (QTL) affecting traits of economic importance. The gene maps or genetic linkage maps of almost all the domestic animals and several model organisms have been completed. Further, comparison of this information with closely related phyletic groups builds the comparative genomics data. Ultimately, the information those are generated in the internet are in the form of database and they include enormous DNA sequence, genetic-, physical- and comparative maps and so on for bioinformatics application.

---

[1] Correspondence: P. Das, Senior Scientist, Central institute of Freshwater Aquaculture; Kausalyaganga, Bhubaneswar2; Email pdas77@hotmail.com

Internet database resource:
http://www.meddb.info/index.php.en?cat=8
http://www.ebi.ac.uk/2can/databases/dna.html
http://www.ebi.ac.uk/2can/databases/genomic.html
http://www.ebi.ac.uk/2can/databases/bib.html

3. Transcriptomics

Expression of a biological trait in an organism is linked to the function of one or combination of few genes. Alternatively, it may be said that genes express the biological traits of humans, animals, plants and microorganisms. The acquisition of new knowledge about genes will make it possible to identify, prevent or treat disease in humans, animals and plants. Using new technology, it is possible to study tens of thousands of genes at once. Thus it has become a realistic objective to determine how the genes in an organism function. At the messenger RNA level, above 5 million ESTs (expressed sequence tags) of the complementary DNAs (cDNAs) of messenger RNAs have been sequenced in humans. In plants, several millions have been sequenced in soyabean, maize, tomato and *Arabidopsis*.. These sequences can be used to create high-density filters, DNA chips or microarrays, where thousands of cDNAs are fixed on a small surface, which is then hybridized with the mRNAs from a given stage or organ. It can thus be answered which genes among thousands are expressed in a particular cell type of an organism, at a particular time and stage. Such transcriptomics tools will soon permit extensive study of transcription and its regulation.

Database Resource:
1. http://www.ncbi.nlm.nih.gov/dbEST
2. http://www.ebi.ac.uk/2can/databases/microarray.html

4. Proteomics

Most genes code for and work through proteins. Proteins may have a wide variety of functions in an organism. Thus it is of great significance to know how proteins function. The complete profile of proteins expressed in a cell is called the cell's proteome. There are thousands of different proteins in each individual cell, and different types of cells contain different sets of proteins. The objective of proteomics is to discover how these proteins function and interact with one another. In order to determine the effect of proteins, it is necessary to know their composition; i.e., the way in which their components are arranged. It is also necessary to know the form that they take, i.e., their three-dimensional structure. Most proteins are coiled in a particular way and can only fulfill their function if they assume the correct three-dimensional structure. Techniques have been developed for visual imaging of proteins, but determining protein structures is technically demanding. Data have been stored for a couple of thousand different protein structures in international databases, but the three-dimensional structure for most of the proteins has not yet been determined. In many cases only parts of these proteins' structures have been determined, perhaps only the biologically active part.

At the protein level, the method used today is still two-dimensional (2D) gel electrophoresis. This separates the denatured proteins according to two independent criteria, the isoelectric point in the first dimension (isoelectrofocusing) and the apparent molecular mass in the second one (electrophoresis in the presence of SDS). On the 2 D gels obtained, several hundreds to several thousands of protein spots (i.e. of gene products) are revealed. These proteins can then be identified or characterized by different methods. Immunological characterization can be done when the identification of a spot is *a priori* suspected or when we are looking for an already known protein on a 2 D gel. The classical protein sequencing of the N-terminal first 10 or 15 amino acids, or of internal sequences, is a very reliable but expensive and relatively slow running method (one spot a day) and requires reasonable amounts of protein. The combination of different criteria can be adopted to identify a protein spot faster and at a lower cost: It permits one to unambiguously identify a protein, if its gene is already present in the databanks. More samples can be analysed for a lower cost than with Edman sequencing. However, protein identifications are performed today by mass spectrometry. The MALDI-TOF (Matrix Assisted Laser Desorption Ionisation - Time of Flight) mass spectrometers give masses of peptides obtained after trypsin digestion of the spots. These high-throughput apparatuses permit a fast identification of numerous proteins when detailed genomic information is available on the studied species. The data obtained ('peptide mass fingerprints') are compared to those generated from the databanks such as SwissProt or 'rEMBL. Another type of mass spectrometer, the ESI-MS/MS (Electro Spray Ionisation - Mass Spectrometer / Mass Spectrometer), gives, by fragmentation of the trypsic peptides, values that are diagnostic of amino acid sequences. Then, as with Edman sequencing, the sequence itself can be compared to homologous ones from other species and is thus preferred when the genome of the studied species is not well represented in databanks.

Internet resource:
http://www.ebi.ac.uk/2can/databases/protein.html
http://www.expasy.ch/enzyme/

5. Metabolomics

Metabolomics is a powerful emerging technology, whereby the total metabolite composition (the metabolome) of an organism is analyzed. By characterizing the metabolome of an organism at different developmental stages, or following exposure to different conditions, global shifts in its metabolism can be followed. This approach complements studies in which changes in the transcriptome and proteome are monitored. By combining global metabolomic analysis with transcriptomics and proteomics it is possible, for the first time, to gain a holistic view of the complex interactions between genes and metabolites. New metabolomic technologies will lead to a greater understanding of metabolism than was possible using previous profiling approaches limited to particular classes of compounds. In addition, metabolomics offers a powerful technology to characterise genes of unknown function, whereby the expression of the gene is manipulated by mutagenesis or genetic engineering and its identity inferred from the resulting metabolic changes. In particular, studies of plants and microbes are important in identifying molecular determinants of food quality, safety and nutrition, and opportunities for the exploitation of novel metabolic products. Plant Molecular Biology (PMB), Biochemistry of Metabolic Regulation in Plants

(BOMRIP), Resource Allocation and Stress in Plants (RASP) and Prokaryotic Responses to Environmental Stress (PRES) are the research sub-disciplines for the generation of data. Over 40,000 biochemical compounds have already been identified in plants. Nevertheless, 10,000s and perhaps even over 100,000 compounds have still to be discovered. These compounds belong to a wide variety of metabolic classes that are all, in one way or another, inter-related. This multitude of chemical components, with many different carbon skeletons, functional groups and physiochemical properties makes plants an invaluable source of pharmaceuticals, health-promoting compounds, flavour & fragrance compounds, protectants, biocides, fine chemicals, toxins, etc.

Internet database resources:
http://www.genome.ad.jp/kegg/metabolism.html
http://www.expasy.ch/cgi-bin/show_thumbnails.pl
http://www.expasy.ch/cgi-bin/show_thumbnails.pl?2

6. Glycomics
Glycomics is the study of interaction between proteins and carbohydrates (glycans). Maccabim-based Glycominds pioneered novel glycomics technologies to analyze interactions between complex carbohydrates and proteins, which will help identify many new drug targets as well as dramatically improve early stage drug development. Now, with the new GlycoChip, the first biochip microarray of glycans, one can analyze unprecedented numbers of glycan-protein interactions. We can see numerous applications of the GlycoChip in the pharmaceutical and biopharmaceutical industry, including identifying novel carbohydrate-binding proteins and potential inhibitors of glycan-protein interactions, as well as analysis of immunogenicity in serum samples for antibody-glycan binding. That is an important step in drug identification because more than 90 percent of therapeutic proteins are bound with glycans critical to the protein function. Combined with Glycominds' Glycomics Database, the world's largest comprehensive database of glycans, includes over 35,000 glycans from various biological sources (1,380 taxons and 205 diseases) describing more than 7,000 unique structures, 1,400 proteins and 1,550 interactions.

Internet database resource:
http://www.meddb.info/index.php.en?cat=8
http://www.heidelberg.de/spec2/sweetdb/start.php?

RECONSTRUCTION OF PHYLOGENETIC TREE USING ONLINE TOOLS

S.N.Rath[1]

All the morden biology is built upon the foundation of evolutionary theory and it is impossible to evaluate relationships among sequences without taking into consideration the way these sequences have been modified over time. Molecular Phylogenetics deals with the study of evolutionary relationships among organisms using molecular data like DNA, Protein sequences, transposable elements and molecular markers. Similarity searches and multiple alignments of sequences naturally lead to the question "How are these sequences related?" and more generally: "How are the organisms from which these sequences come related?". Sequence analysis of biological data is based on some evolutionary principles which state that group of similar organisms descended from a common ancestor.

Both nucleotide data and amino acid data can be used to generate phylogeny and there has been much debate about which is the best. The main argument of using amino acid data to infer phylogeny that there are 20 characters in comparison to 4 characters and also the alignment of amino acid sequence data is easier than DNA sequence data. A umber of evidences, which supports the reconstruction of phylogeny from protein sequences, are very much useful than DNA sequence [1].

- Codon bias
- Long time Horizon
- Advantages of translation of DNA into corresponding protein sequences
- Nature of sequence divergence in protein
- Proteins have different rate of evolution
- Introns and the noncoding DNA
- Multigene families
- Protein is the unit of selection
- RNA editing

But it is advisable to analyse both data sets (DNA and protein) in favorable condition.

- As for a group of species or taxa that are relatively close in time or that are closely related (like viral proteins or vertebrate enzymes) DNA-based analysis is probably a good way to go, since problems as differences in codon bias or saturation of the third position of codons can be avoidable in this case. It is nevertheless strongly recommended to carry out an analysis on the protein data as well.

[1] Correspondence: Surjya Narayan Rath, *Asst.Professor;* Dept. Bioinformatics, CPGS, Orissa University of Agriculture & Technology; Bhubaneswar-751003; snr_bioinfo@rediffmail.com

- Multigene families (for instance genes coding for different, but similar, isoenzymes) may cause problems and it needs careful decision to exclude or include such sequences in data set and peculiarly looking phylogenetic trees [2].

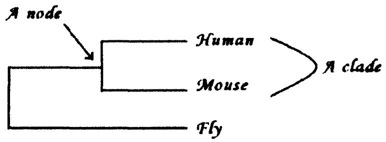
Some terms Associated with Phylogeny

Phylogenetic tree: A graph that illustrates the evolutionary relationships using nodes and branches.

Node: It represents a taxonomic unit. This can be either an existing species or ancestors.

Branch: It defines relationship among the in term of descent and ancestry.

Branch length: It represents the number of changes that have occurred in the branch.

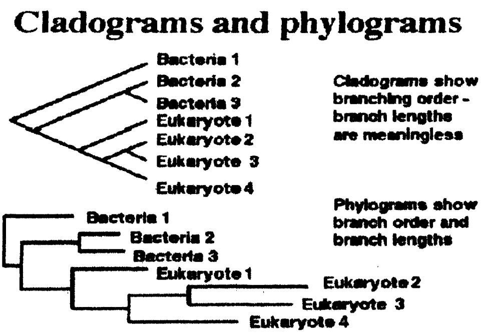


(Fig-1)

Methods of Phylogenetic analysis

There are two major groups of analysis for studying phylogenetic relationships.

- Phenetic methods
- Cladistic methods



(Fig-2)

There are a lots of online tools for phylogenetic tree construction , but among them Phylip, Paup and MEGA are most widely used due to its user-friendly ness and reliability.

Steps to reconstruct a tree:

- Collection of gene or protein sequences of different species of your interest.
- Input format preparation for multiple sequence analysis.
- Multiple sequence alignment by using tools like Clustal W, Clustal X, Pileup of the Wisconsin GCG (Genetic Computer Group) package.
- Tree construction by running any phylogeny software like Phylip 3.67, PAUP* 4.0 or MEGA-4.

References

http://www.icp.be/~opperd/private/arguments.html#anchor574630
http://www.icp.be/~opperd/private/proteins.html#anchor2815051
http://paup.csit.fsu.edu/
http://evolution.genetics.washington.edu/phylip.html
http://www.megasoftware.net/