

PROCEEDINGS

International Conference on ENVIRONMENTAL AND MEDICAL STATISTICS

9-10 January 2020

Editorial Board

Sachith Abeysundara (Editor-in-Chief)

Pushpakanthie Wijekoon

Namal Priyantha

Inoka Karunarathne

Lakshika Nawarathna

Organized by the

**DEPARTMENT OF STATISTICS AND COMPUTER SCIENCE,
FACULTY OF SCIENCE AND
POSTGRADUATE INSTITUTE OF SCIENCE (PGIS)
UNIVERSITY OF PERADENIYA, SRI LANKA**

**DEPARTMENT OF MATHEMATICS & STATISTICS,
UNIVERSITY OF MARYLAND BALTIMORE COUNTY, USA**

Copyright © 2020 by
Department of Statistics and Computer Science,
University of Peradeniya

All rights reserved. No part of this publication may be reproduced, distributed, stored in a retrieval system, and transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

Published by
Department of Statistics and Computer Science,
Faculty of Science,
University of Peradeniya,
Peradeniya 20400,
SRI LANKA.

Printed by
Sanduni Offset Printers (Pvt) Ltd,
4/1, Sarasavi Uyana Goodshed Road,
Sarasavi Uyana, Peradeniya 20400,
SRI LANKA.

Cover design: Mr. P.M.P.C. Gunathilake

Printed in the Democratic Socialist Republic of Sri Lanka.

TABLE OF CONTENTS

Message from the Dean, Faculty of Science	v
Message from the Director, Postgraduate Institute of Science	vii
Messages from the Conference Co-chairpersons	ix-xi
List of Abstracts	xiii
Abstracts	
Plenary speeches	1
Keynote speeches	3
Invited speeches	10
Theme-I : Environmental Statistics.....	21
Theme-II : Medical Statistics	41
Conference Organizing Committee.....	57
Sponsors	59
Author Index	61

ICEMS 2020

Message from the Dean Faculty of Science, University of Peradeniya, Sri Lanka

It is indeed my pleasure to welcome all of you to the inaugural International Conference on Environmental and Medical Statistics jointly organized by the Faculty of Science, Postgraduate Institute of Science (PGIS), University of Peradeniya and the University of Maryland Baltimore County, USA. The Conference is organized as a set of themes in Environmental and Medical Statistics and it provides an opportunity for researchers to interact themselves.

I am honoured to have Dr. Barry Nussbaum and Professor Thomas Mathew as distinguished plenary speakers of the Conference and several keynote speakers who are experts in fields of Environmental and Medical Statistics. I hope that this Conference will create a platform for researchers to discuss their research findings and the deliberations of the research findings which will prove productive and intellectually stimulating discussions and dialogues.

I hope that you will find the Conference both valuable and enjoyable, and also you will find free time to enjoy the architectural, cultural and natural beauty of University of Peradeniya and historical cultural heritage city of Kandy and the other parts of Sri Lanka.

It is important to record that the successful organization of the Conference required the talents, dedication and time of many volunteers and strong support from sponsors. Without their outstanding contribution and nearly a year of planning, International Conference on Environment and Medical Statistics would not have been realistic.

I wish International Conference on Environmental and Medical Statistics will be a great success!

Professor Saluka R. Kodithuwakku

*Dean, Faculty of Science,
University of Peradeniya.*

ICEMS 2020

Message from the Director Postgraduate Institute of Science, University of Peradeniya, Sri Lanka

It is my great pleasure and honour to welcome you on behalf of the Postgraduate Institute of Science (PGIS) to the International Conference on Environmental and Medical Statistics (ICEMS) organized by the Department of Statistics and Computer Science, Faculty of Science and the Postgraduate Institute of Science University of Peradeniya jointly with Department of Mathematics & Statistics, University of Maryland Baltimore County, USA. We are extremely grateful to the two Co-chairpersons, Professor Bimal Sinha and Professor Pushpa Wijekoon, and the organizing committee for the time and effort they put into arranging it.

Statistical analysis allows studying the climate, environment and medical sciences which includes the nature and impact of extreme events, diseases, patients, and epidemiological events. The modern researcher cannot refuse to know and to use statistics. This conference will share an insight into the recent developments in statistical methodologies in environmental and medical sciences, which gains immense interest with the presence of academic experts, young and brilliant researchers, and talented student communities.

I believe that this conference will promote top level research and to popularize the quality research in general, thus making discussions, presentations focusing attention on the recent developments in the area, and future trends and needs.

My sincere thanks go to the many reviewers and the Committee members who worked hard to ensure the quality of the selected proceedings.

Please accept my warmest wishes for an enjoyable event at the PGIS.

Professor H.M.T.G.A. Pitawala

*Director, Postgraduate Institute of Science,
University of Peradeniya.*

ICEMS 2020

Message from the Congress Co-chairperson

It is a great pleasure and privilege for me to welcome you on behalf of the Department of Statistics & Computer Science, Faculty of Science and the Postgraduate Institute of Science (PGIS), to the first International Conference on Environmental and Medical Statistics (ICEMS). We are honoured to have the Department of Mathematics and Statistics, University of Maryland Baltimore County, USA as a collaborative partner for this conference.

In the Environmental and Medical Sciences, statistics plays a key role in planning the researches, and making predictions and decisions about the future management of the relevant scenario. This conference is an outstanding opportunity to bring forth recent developments in the field of Environmental and Medical statistics and related areas, and to bring together leading academics, scientists, research scholars and research students to exchange and share their experiences and research findings. It also provides a platform to present and discuss the most recent innovations, trends, and concerns as well as practical challenges encountered, and solutions adopted in the fields of Environmental and Medical Statistics.

We are honoured to have Mr. Dhirendra Singh, Assistant High Commissioner of India as the Chief Guest, and Prof. Upul B. Dissanayake, the Vice Chancellor, University of Peradeniya, as the Guest of Honour. It is a great pleasure to have the two distinguished plenary speakers Dr. Barry Nussbaum, a past president of American Statistical Association and a former Chief Statistician of Environment Protection agency, USA, and Prof. Thomas Mathew, Department of Mathematics and Statistics, University of Maryland Baltimore County, USA. Further, we are proud to have seven eminent researchers as keynote speakers, six from universities of USA, and one from an Australian university. The conference is also enriched with 6 invited speeches of outstanding researchers, three from universities of USA, and three from local universities. In addition to the plenary, keynote, and invited talks, the ICEMS is also rich and varied with around 35 contributed papers from various national and international universities, research institutions and industries.

The success of the conference ultimately depends on the talents, dedication and time of many academic and non academic staff including the volunteers. Our special gratitude and appreciation are to the Co-chairperson, Prof. Bimal Sinha from the Department of Mathematics and Statistics, University of Maryland Baltimore County, USA for his wisdom and guidance throughout the organizing period. Recognition should also go to the local organizing committee members who have worked extremely hard for the success of this event, and to the reviewers for their thorough and timely reviewing of the papers. Further, the advice and the brilliant support given by the Director/PGIS, Dean, Faculty of Science, Vice Chancellor and Deputy Vice Chancellor of University of Peradeniya, and the services rendered by the PGIS staff and other non academic branches of the university are greatly appreciated. We also would like to thank all of the sponsoring organizations for providing their generous financial support, and all of the conference presenters for their contributions which are the foundation of this conference. Finally, I wish all the participants a highly stimulating and excellent conference for knowledge exchange, insights and collaborations.

Professor Pushpakanthie Wijekoon

Co-chairperson

International Conference on Environmental and Medical Statistics

ICEMS 2020

Message from the Congress Co-chairperson

I am delighted to welcome all of you to the First International Conference on Environmental and Medical Statistics (ICEMS) to be held during January 9-10, 2020, under the joint collaboration of the Department of Statistics & Computer Science, Faculty of Science and the Postgraduate Institute of Science (PGIS, University of Peradeniya and the Department of Mathematics and Statistics, University of Maryland Baltimore County, USA.

Novel applications of statistical methods in Environmental and Medical Sciences have resulted in many serious significant developments, leading to better management of health and environmental factors. The conference provides an ideal forum to bring together distinguished scientists from both sides of the Atlantic to this beautiful island nation of Sri Lanka, and I sincerely hope that there will be plenty of discussions on the subject matters among the conference participants for mutual benefit.

I am thankful to Mr. Dhirendra Singh, Assistant High Commissioner of India and Prof. Upul B. Dissanayake, Vice Chancellor of the University of Peradeniya to address the conference attendees with their insightful comments. I am grateful to all the domestic and international speakers, including my distinguished UMBC colleagues Professor Nagaraj Neerchal and Professor Thomas Mathew. I believe their scientific deliberations will enrich the topics of the conference, and immensely benefit the young researchers.

I thank the members of the local organizing committee for their hard work, making sure the conference proceeds smoothly. I have no words to express my gratitude to my wonderful friend Professor Pushpa Wjekoon, Co-Chairperson of this conference, for her overall supervision of every aspect of this conference.

I wish this exciting international conference a grand success!

Professor Bimal K. Sinha

Co-chairperson

International Conference on Environmental and Medical Statistics

ICEMS 2020

PROCEEDINGS

List of Abstracts

Plenary Speeches

Abstract ID	Title and Authors	Page No.
PS1	NEW CHALLENGES FOR TODAY'S STATISTICIAN: HOW THEY AFFECT ENVIRONMENT AND HEALTH <i>Barry D. Nussbaum</i>	01
PS2	INFERENCE ON THE DRUG INTERACTION INDEX BASED ON BINARY RESPONSE DATA <i>Thomas Mathew</i>	02

Keynote Speeches

Abstract ID	Title and Authors	Page No.
KS1	STATISTICAL METHODS FOR PRIVACY AND DATA CONFIDENTIALITY PROTECTION IN MEDICAL RESEARCH <i>Tapan Nayak</i>	03
KS2	PREDICTION METHODS FOR SEMI-CONTINUOUS DATA WITH APPLICATIONS IN CLIMATE SCIENCE <i>Nagaraj K. Neerchal</i>	04
KS3	EVIDENCE OF A LINEAR RELATIONSHIP BETWEEN VALENCE AND AROUSAL FOR SUBJECTIVE EXPERIENCE IN A REAL-LIFE SETTING USING LONGITUDINAL ANALYSIS OF ECOLOGICAL MOMENTARY ASSESSMENT (EMA) DATA <i>Rajesh Ranjan Nandy</i>	05
KS4	AIR QUALITY MONITORING ANALYSIS OF AIRBORNE MICROBES USING PORTABLE LOW-COST DEVICE <i>Shantanu Sur</i>	06
KS5	A SEGMENTED MEASUREMENT ERROR MODEL FOR MODELING AND ANALYSIS OF METHOD COMPARISON DATA <i>Pankaj K. Choudhary</i>	07
KS6	SPATIO TEMPORAL DATA INTEGRATION PLATFORM FOR EVALUATING ENVIRONMENTAL IMPACTS ON PUBLIC HEALTH <i>Liwan Liyanage</i>	08
KS7	TOPIC MODELING FOR IDENTIFYING VARIATION IN MICROBIAL COMMUNITIES IN PLANTS AND SOIL ENVIRONMENT <i>Pratheepa Jeganathan</i>	09

Invited Speeches

Abstract ID	Title and Authors	Page No.
IS1	FEATURE SELECTION OF RHEUMATOID ARTHRITIS COMORBIDITIES FOR DISEASE PREDICTION <i>Sumona Mondal</i>	10
IS2	DETECTING RARE HAPLOTYPE ASSOCIATION WITH TWO CORRELATED PHENOTYPES OF BINARY AND CONTINUOUS TYPES <i>Swati Biswas</i>	11
IS3	SMALL AREA ESTIMATION OF PREVALENCE USING HIERARCHICAL AND EMPIRICAL BAYES METHODS WITH APPLICATIONS TO UNINSURED MINORITY SUB-POPULATION AND TOBACCO USE AMONG NURSES <i>Karabi Nandy</i>	12
IS4	GOODNESS OF FIT FOR A JOINT MODEL PERTAINING TO MEDICAL DATA <i>M.R. Sooriyarachchi and Nayani Adikari</i>	13
IS5	JOINT MODELING OF DENGUE SURVIVAL AND COUNTS WITH JOINT GENERALIZED ESTIMATING EQUATIONS <i>J.C. Hapugoda and M.R. Sooriyarachchi</i>	14
IS6	COMPARISON OF JOINT MODELING APPROACHES FOR ANALYZING CORRELATED DISCRETE COMPETING RISK AND CONTINUOUS OUTCOME <i>G.H.S. Karunaratna and M.R. Sooriyarachchi</i>	15
IS7	CAN CLIMATE FACTORS INFLUENCE CKD OF SRI LANKA? <i>K.W.G. Rekha Nianthi and Dhammadika Dissanayake</i>	16
IS8	NON-METRIC TOOTH CROWN TRAITS IN A CONTEMPORARY AND TWO ABORIGINAL POPULATIONS IN SRI LANKA: COMPARISON WITH OTHER WORLD POPULATIONS <i>H.R.D. Peiris, J.A.A. Jayaweera, A.K.S. Arambawatta and C.D. Nanayakkara</i>	17

THEME-I : Environmental Statistics

Abstract ID No.	Title and Authors	Page No.
01	IMPACT OF ENERGY CONSUMPTION AND ECONOMIC GROWTH ON CARBON DIOXIDE EMISSIONS IN SRI LANKA <i>K.A.U. Sakunthala and T.S.G. Peiris</i>	21
04	UNDERSTANDING CHRONIC KIDNEY DISEASE IN HISTORICALLY GROUNDWATER RELIANT COMMUNITIES IN SRI LANKA: A SOCIOECONOMIC APPROACH <i>K. Kafle, S. Balasubramanya, T. Horbulyk, D. Stifel and H. Manthrithilake</i>	22
05	GROUNDWATER QUALITY AND ALTERNATIVE SOURCES OF GROUNDWATER IN MOST CHRONIC KIDNEY DISEASE AFFECTED COMMUNITIES IN SRI LANKA <i>S. Balasubramanya, T. Horbulyk, K. Kafle, D. Stifel and H. Manthrithilake</i>	23
07	FITTING A GENERALIZED LINEAR MODEL FOR RAINFALL OCCURRENCE AND RAINFALL AMOUNT IN COLOMBO DISTRICT <i>A. Selvaratnam and P. Wijekoon</i>	24

Abstract ID No.	Title and Authors	Page No.
08	PREDICTING ANNUAL STEM DIAMETER INCREMENT OF SELECTED TREE SPECIES IN SINHARAJA RAIN FOREST CONSIDERING TREE AND STAND LEVEL EFFECTS <i>D.P.V.R. Dissanayake and P. Wijekoon</i>	25
12	THERMAL COMFORT AND CLIMATE REGIONALIZATION IN SRI LANKA <i>H.M.N.S. Kumari and P. Wijekoon</i>	26
13	DEVELOPMENT OF LONG MEMORY MODEL TO FORECAST WEEKLY RAINFALL <i>H.P.T.N. Silva and T.S.G. Peiris</i>	27
14	PUBLIC ACCEPTABILITY OF TREATED SURFACE WATER FOR POTABLE PURPOSE: CASE STUDY IN VAVUNIYA <i>M.A.O.D. Mathara Arachchi and S. Devaisy</i>	28
16	ANALYSIS OF ASSOCIATION BETWEEN DENGUE CASES AND RAINFALL IN SRI LANKA <i>S.M.J.H. Nisansala and P. Wijekoon</i>	29
18	DATA EXPLORATION AND PRE-PROCESSING TECHNIQUES ON AIR POLLUTION AND METEOROLOGICAL DATA IN SYDNEY REGION <i>W. M. L. K. N. Wijesekara and L. Liyanage</i>	30
19	WASTE-DERIVED BIOSORBENTS TO REMOVE ORGANICS FROM RICE MILL WASTEWATER <i>U.D.T.M. Uduwarage and S. Devaisy</i>	31
24	SHORT TERM ELECTRICITY DEMAND FORECASTING FOR GENERATION SCHEDULING: A CASE STUDY IN SRI LANKA <i>E.D.T. Somaratne, D.S.K. Karunasinghe and J.V. Wijayakulasooriya</i>	32
25	SPATIAL AND TEMPORAL VARIATION OF LAND SURFACE TEMPERATURE IN COLOMBO REGION <i>R.P.G. Rajapaksha, K.W.G. Rekha Nianthi and S.I.S. Subasinghe</i>	33
27	DETERMINATION OF PHOTOSYNTHETIC BIOMASS, OXYGEN PRODUCTION, AND WATER PRODUCTION IN THE EARLY GROWTH STAGES OF AGARWOOD <i>H.N. Chathuranga, F.R. Senanayake, M.K.L Chandana and D. De Zoysa</i>	34
33	BANANA PSEUDO STEM BIOMASS AS A PROSPECTIVE ADSORBENT TO REMOVE INDUSTRIAL DYES <i>B.K. Pathiraja and U.K. Jayasundara</i>	35
39	A STUDY OF VARIATIONS IN THE RAINFALL OF THE WESTERN PROVINCE BASED ON REPRESENTATIVE CONCENTRATION PATHWAY SCENARIOS <i>L. K. C. Tillekeratne, W. C. W. Navaratna and G. D. P. Dharmarathna</i>	36
40	USING BLACK-BOX MACHINE LEARNING TECHNIQUES TO IDENTIFY SPATIAL DEPENDENCE IN OCCUPANCY DATA <i>N.M. Mohankumar and T.J. Hefley</i>	37

THEME-II : Medical Statistics

Abstract ID No.	Title and Authors	Page No.
10	TRENDS OF SEXUAL ABUSE IN SRI LANKA <i>R.A.S.U Ranasinghe, S.D. Dharmaratne, L.I Ranasinghe, I.N Thotawatte, P.G.K De Silva and W.D.T.S Welwita</i>	41
11	MODELING THE DENGUE EPIDEMIOLOGY IN SRI LANKA USING A RANDOM EFFECT COPULA MODEL <i>G.H.S. Karunaratna and M.R. Sooriyarachchi</i>	42
15	ASSESSMENT OF CORRELATION BETWEEN MAMMOGRAPHIC BREAST DENSITY AND FACTORS AFFECTING BREAST DENSITY: A QUANTITATIVE APPROACH <i>I.L.U. Chandrasiri, P.B. Hewavithana, U.J.M.A.L. Jayasinghe, P.M.P.C. Gunathilake and S.P. Abeysundara</i>	43
17	IDENTIFYING PRIMARY FACTORS THAT AFFECT THE CHRONIC KIDNEY DISEASE WITH A COMPARATIVE ANALYSIS OF HEAVY METALS IN DRINKING WATER <i>P.M.S.S. Kumari and P. Wijekoon</i>	44
20	ASSESSMENT OF OSTEOPOROSIS RISK USING PREDICTIVE MODELS <i>B.G.G.K. Somathilake, S.D. Viswakula and J.M.K.C. Jeewandara</i>	45
22	MODEL-BASED APPROACH TO OPTIMIZE CLINICAL OUTCOMES IN NEONATAL ABSTINENCE SYNDROME USING REAL WORLD DATA <i>N. Wijekoon, O. Aduroja, J. M. Biggs, D. El-Metwally and M. Gopalakrishnan</i>	46
26	PREDICTION OF AGE FROM ERUPTION STATUS OF PERMANENT TEETH IN SRI LANKAN CHILDREN <i>B.D.H.H. De Silva, L. S. Nawarathna and V.S.N. Vithanaarachchi</i>	47
29	MULTI-CLASS CLASSIFICATION APPROACH TO IDENTIFY THE STAGE OF AUTISM SPECTRUM DISORDER <i>M. Kayanan and P. Wijekoon</i>	48
30	EARLY RECOGNITION OF PREMATURE BIRTH: A COMPARATIVE ANALYSIS <i>T.M.S. Priyankara and S.P. Abeysundara</i>	49
31	SERUM TGF BETA 1 AS AN EARLY MARKER OF CHRONIC KIDNEY DISEASE OF UNCERTAIN AETIOLOGY <i>W.B.N.T. Fernando, K.R.D. Hemage, M.Z. Badurdeen, T.W. Hettiarachchi, H.T.K. Abeysundara and N. Nanayakkara</i>	50
35	SEX DETERMINATION USING THE HUMERUS OF SKELETONS OF CONTEMPORARY SRI LANKAN POPULATION <i>T.Selvamalai, L.S.Nawarathna and D. Nanayakkara</i>	51
37	DETERMINATION OF SEX USING FOOT, HAND AND HAND PRINT MEASUREMENTS IN A SRI LANKAN POPULATION <i>A.B.S.N. Dayarathne, L. S. Nawarathna and D. Nanayakkara</i>	52

Abstract ID No.	Title and Authors	Page No.
42	PREVALENCE OF HYPERURICEMIA AND ANAEMIA AMONG PATIENTS WITH CHRONIC KIDNEY DISEASE OF UNKNOWN AETIOLOGY IN WILGAMUWA, SRI LANKA <i>H.P.M. Hewavitharane, N. Nanayakkara, S. Anand, R. Chandrajith, P. Vlahos, S. Schensul, J. Schensul, L. Haider and K.T. Silva</i>	53
43	TRACE ELEMENTS LEVELS IN RICE CONSUMED BY PATIENTS WITH TUBULITIS <i>O.S.S. Preamrathne, R. Chandrajith, N. Nanayakkara, H.P.M. Hewavitharane, S. Wijethunge, R.D. Yapa, H.T.K. Abeysundara, N. Elladeniya, D.P.G.S. Kumudumali, K.P.S. Madushan, K.O.C.U. Samarasiri</i>	54
44	PARTICIPATION OF SPORT IMPROVING ACADEMIC ACHIEVEMENTS, HEALTHY LIVING AND WELL-BEING <i>L. Liyanage and R.P. Premaraja</i>	55
45	LATENT RELATIONSHIPS AMONG DEMOGRAPHIC, SOCIO-ECONOMIC, ANTHROPOMETRIC AND SERUM BIOCHEMICAL CHARACTERISTICS IN A POPULATION ON DIETARY ADVICE: MULTIPLE CORRESPONDENCE ANALYSIS APPROACH <i>G.G.S. Ranasinghe and S. Samita</i>	56

Plenary Speech (PS1)

NEW CHALLENGES FOR TODAY'S STATISTICIAN: HOW THEY AFFECT ENVIRONMENT AND HEALTH

Barry D. Nussbaum

*Department of Mathematics and Statistics, University of Maryland Baltimore County,
Baltimore, MD 21250, USA.
statisticsbarry@gmail.com*

Statisticians must be ever so vigilant in presenting results that reflect useful, practical, innovative, and current analyses. These must be succinct and understandable to be used for important decisions regarding medical and environmental problems. But today we have challenges regarding “big data”, significance debates, misleading breakthroughs, public understanding, and disparate sample sizes. It is imperative that we address these concerns. This presentation will explore these challenges with some important examples.

Keywords: Big Data, Public Understanding, Sample Size, Significance, Simpson’s Paradox.

Plenary Speech (PS2)

**INFERENCE ON THE DRUG INTERACTION INDEX BASED ON
BINARY RESPONSE DATA**

Thomas Mathew

*Department of Mathematics and Statistics, University of Maryland Baltimore County,
Baltimore, MD 21250, USA.
mathew@umbc.edu*

The interaction index is a parameter that is used to quantify the interaction when two or more drugs are combined in a combination therapy. The index is defined based on the deviation of the response of the drug combination, from the expected response obtained under a reference model of no interaction. The well-known Loewe additivity model is often used to derive the interaction index; when the drugs do not interact, the interaction index takes the value one. The drugs act synergistically when the index is less than one, and antagonistically when the index is more than one. In the talk, I will consider binary response data and use the fiducial approach to derive a confidence interval for the interaction index. Accuracy of the proposed solution will be assessed based on estimated coverage probabilities, and will be compared with available large sample solutions based on the expected width of the confidence interval. The results will be illustrated using data from a study to assess interaction in a combination therapy involving the analgesic drugs Tramadol and Acetaminophen.

Keywords: Drug Interaction, Binary Response Data, Loewe Additivity model, Analgesic drugs.

Keynote Speech (KS1)

**STATISTICAL METHODS FOR PRIVACY AND DATA CONFIDENTIALITY
PROTECTION IN MEDICAL RESEARCH**

Tapan Nayak

*Department of Statistics, George Washington University, Washington, DC 20052, USA.
tapan@gwu.edu*

Recent advances in big data analytics can be applied to medical records from multiple sources, e.g., hospitals, doctor's offices and clinical studies, to develop new drugs, improve treatments, detect epidemics, reduce healthcare costs and derive other benefits. However, privacy concerns and regulations are major obstacles to sharing patient records. Removing direct identifiers such as name and social security number is inadequate as sensitive information about a patient might be gained by matching some easily available demographic variables such as gender, age and race. Thus, only a suitably masked or perturbed version of the actual data can be released. Naturally, masking also distorts or dilutes statistical information. The challenge is to develop data masking methods that protect privacy without losing much data utility. We shall describe some important privacy concerns and present some statistical ideas and methods for privacy protection.

Keywords: Privacy Protection, Data Masking methods, Data Utility.

Keynote Speech (KS2)

**PREDICTION METHODS FOR SEMI-CONTINUOUS DATA WITH
APPLICATIONS IN CLIMATE SCIENCE**

Nagaraj K. Neerchal

*¹Department of Mathematics and Statistics, University of Maryland Baltimore County,
Baltimore, MD 21250, USA.
nagaraj@math.umbc.edu*

Semi-continuous random variables have discrete and continuous components with support on a set of discrete points and a subset on the real line. Daily precipitation (rainfall) data is an example of such a random variable with a point mass at 0 and an absolutely continuous distribution function on the positive real line. When the Probability of observing a 0 is assumed to be independent of the parameters of the continuous part, the density of the random variable takes the form a Two-Part model. A popular form that enforces a dependency is the standard Tobit model. We briefly review some inferential aspects of the semi-continuous distributions and present several methods of prediction and derivation of predictive densities, motivated by applications of spatio-temporal models in Climate Science.

*This is a joint work with Chinmaya Vishwavidyapeeth, Adi Shankara Nilayam, Adi Shankara Marg,
Veliyanadu, Kerala 682313, India.*

Keywords: Semi-continuous data, Two-part model, Tobit model, Spatio-temporal models.

Keynote Speech (KS3)

EVIDENCE OF A LINEAR RELATIONSHIP BETWEEN VALENCE AND AROUSAL FOR SUBJECTIVE EXPERIENCE IN A REAL-LIFE SETTING USING LONGITUDINAL ANALYSIS OF ECOLOGICAL MOMENTARY ASSESSMENT (EMA) DATA

Rajesh Ranjan Nandy

*Department of Biostatistics and Epidemiology, University of North Texas Health Science Center,
Fort Worth, TX 76107, USA.
rajesh.nandy@unthsc.edu*

The circumplex model of affect posits that valence and arousal are the principal dimensions of affect. However, no consistent relationship between valence and arousal has been established. Ecological momentary assessment (EMA) data on affect from subjects residing in permanent supportive housing (PSH) was used to study the relationship between valence and arousal in real-life settings. Longitudinal analysis of the EMA data reveals a strong positive linear relationship at both nomothetic and idiographic level, which is the first finding of its kind. These findings open up the possibility that there may be a universal structure of affect.

Keywords: Circumplex model, Ecological Momentry Assessment, Permenant Supportive housing, Logitudinal analysis.

Keynote Speech (KS4)

AIR QUALITY MONITORING ANALYSIS OF AIRBORNE MICROBES USING PORTABLE LOW-COST DEVICE

Shantanu Sur

Department of Biology, Clarkson University, Potsdam, NY 13699, USA.
ssur@clarkson.edu

Information regarding the pathogenic microbes in the ambient air can be highly beneficial to contain the outbreak of airborne transmissible diseases or reduce healthcare-associated infections. To address such need, in conjunction with a local startup, we have developed a portable, low-cost, low-power device that is capable of real-time air quality (AQ) monitoring along with the capture of airborne particles. The device named as IoT Real-time Aerosol Concentration sensor and Biological sampler (TRAC-B) uses an optical sensor to measure aerosol concentration and the data is transferred to cloud in real-time. An ionizer and an electrostatic precipitator integrated to the device enable the capture of bioaerosol sample on a removable collection plate. Collected samples are analyzed by a combination of next-generation sequencing (NGS), quantitative PCR, and colony culture followed by Sanger sequencing to estimate the microbial abundance and diversity. I will discuss some useful techniques to visualize and analyze these data with examples of sample data collected from various indoor and outdoor locations. Our eventual goal is to deploy multiple sensor devices distributed across the target location to generate heatmaps of sensor network-derived AQ and airborne microbe data and implement data analytics tools to predict pathogen risks from recognizable patterns in AQ at a high degree of spatiotemporal precision.

Keywords: Real-time monitoring, Low-cost sensing, Bioaerosol sampling, Microbial diversity.

Keynote Speech (KS5)

A SEGMENTED MEASUREMENT ERROR MODEL FOR MODELING AND ANALYSIS OF METHOD COMPARISON DATA

Pankaj K. Choudhary

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080, USA.
pankaj@utdallas.edu

Method comparison studies are concerned with estimating relationship between two clinical measurement methods. The methods often exhibit a structural change in the relationship over the measurement range. Ignoring this change would lead to an inaccurate estimate of the relationship. Motivated by a study of two digoxin assays where such a change occurs, this work develops a statistical methodology for appropriately analyzing such studies. Specifically, it proposes a segmented extension of the classical measurement error model to allow a piecewise linear relationship between the methods. The changepoint at which the transition takes place is treated as an unknown parameter in the model. An expectation-maximization type algorithm is developed to fit the model and appropriate extensions of existing measures are proposed for segment-specific evaluation of similarity and agreement. Bootstrapping and large-sample theory of maximum likelihood estimators are employed to perform the relevant inferences. The proposed methodology is evaluated by simulation and is illustrated by analyzing the digoxin data.

This is a joint work with L.N.K. Rankothgedara, Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080, USA.

Keywords: Agreement, Changepoint, Concordance Correlation, ECM Algorithm, Total Deviation Index.

Keynote Speech (KS6)

SPATIO TEMPORAL DATA INTEGRATION PLATFORM FOR EVALUATING ENVIRONMENTAL IMPACTS ON PUBLIC HEALTH

Liwan Liyanage

*School of Computing, Engineering and Mathematics, Western Sydney University,
Locked Bag 1797, Penrith NSW 2751, Australia
l.liyanage@westernsydney.edu.au*

This multidisciplinary project is designed to measure the impacts of environmental conditions and changes on public health, to obtain vulnerability profiles, then to translate to policies for managing and mitigating burden of disease. This data integration platform is a big data source for evaluating environmental (climate, pollution and demographic) impacts on many relevant application domains in spatio-temporal scale. Such applications include and not limited to health and disease. Primarily considering as the core elements the climate and pollution data is uploaded and integrated via position and time variables using nearest neighbour approach developing a spatiotemporal data source for the state Victoria within Australia. All the tools used are open source. Data repository was implemented using PostgreSQL and PostGIS, which supports geo-referencing and storage of large data. The spatio-temporal visualization was done using QGIS application as it integrates with PostgreSQL and PostGIS. Statistical analysis, spatial analysis was done using R statistical software and QGIS. Data custodians are Environmental Protection Authority, Bureau of Meteorology, Bureau of Statistics and Local Health Authority Geelong Hospital. This provides the opportunity to create data driven methods to conduct a comprehensive study on effects and interactions with cross-domain variables. Further data visualization tool visually assesses the impacts on varying time scales and on varying locations. Analysis of this integrated data source provide ample opportunities to find multi-factorial patterns related to social and environmental (weather, pollution and demographic) conditions on public health. Hence guide the policy makers to look at vitally important health issues at multidimensional scale.

Keywords: Data Integration, Environmental Impact, Open Source, Public Health, Spatio Temporal Data.

Keynote Speech (KS7)

TOPIC MODELING FOR IDENTIFYING VARIATION IN MICROBIAL COMMUNITIES IN PLANTS AND SOIL ENVIRONMENT

Pratheepa Jeganathan

Department of Statistics, Stanford University, Stanford, CA 94305, USA.
pjeganat@stanford.edu

Next-Generation Sequencing (NGS)-based technology now enables the high-throughput quantification of non-culturable microbial organisms in all environments. The use of microbial sequences has enhanced our understanding of the microbiome interactions at the interface of host-plant and soil environment and their roles in microbial ecology. These implications can be used in the plant microbiome-based agro-management to improve agricultural production, promote plant growth and health, keep resistance against diseases, and environmental stress. Nevertheless, the plant microbiome data pose statistical challenges such as 1) amplification of host contaminations such as plastid and mitochondria, 2) heterogeneous microbiome communities in host-plants, and soil environment, 3) sequencing of unequal sampling depth, 4) latent microbial communities that involve in interactions. We propose a Latent Dirichlet Allocation to identify microbial communities in plant root rhizosphere and endosphere and their interactions at the interface to exchange resources. We illustrate our method by reusing the 16S rRNA gene sequencing data of the rhizosphere and endosphere in nine different plant species with the approach to limit the co-amplification of host-plant contamination across different plant species.

Keywords: Bayesian Analysis, Environmental Statistics, Latent Dirichlet Allocation, Microbial Ecology, Next-Generation Sequencing.

Invited Speech (IS1)

**FEATURE SELECTION OF RHEUMATOID ARTHRITIS COMORBIDITIES
FOR DISEASE PREDICTION**

Leon Lufkin¹, Marko Budisic¹, Sumona Mondal^{1,*} and Shantanu Sur²

¹*Department of Mathematics, Clarkson University, Potsdam, NY 13699, USA.*

²*Department of Biology, Clarkson University, Potsdam, NY 13699, USA.*

**smondal@clarkson.edu*

Rheumatoid Arthritis (RA) is a chronic and debilitating auto-immune disorder affecting joints and internal organs that requires diagnosis and management under rheumatologist supervision. With a complex repertoire of effects on multiple body systems, patients suffering from RA are predisposed to several comorbidities such as depression and obesity. However, RA patients living in rural areas experience many barriers to proper care due to a severe shortage of rheumatologists. One potential strategy to improve the care for rural RA is to identify the critical factors in diagnosing and predicting the disease. In this work, we use Bayesian logistic regression to model the relationship between RA and its common comorbidities such as high Body Mass Index, depression, and systolic blood pressure. Using data from National Health and Nutrition Examination Survey for the model, our goal is to accurately quantify the uncertainty to make meaningful inferences. We use wrapper feature selection methods for their high accuracy in identifying the important comorbidities for predicting the disease in individuals. We use the area under the receiver operating characteristic curve and average precision to measure and maximize discrimination ability. Model classification is optimized using F_β and balanced accuracy metrics to appropriately balance failures to identify and false diagnoses of RA and maximize the beneficial impact of our project. The results of this analysis will be discussed in our presentation.

Keywords: Feature selection, Bayesian, comorbidities, prediction, Rheumatoid Arthritis, NHANES.

Invited Speech (IS2)

DETECTING RARE HAPLOTYPE ASSOCIATION WITH TWO CORRELATED PHENOTYPES OF BINARY AND CONTINUOUS TYPES

Xiaochen Yuan and Swati Biswas^{*}

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX 75080, USA.
* *swati.biswas@utdallas.edu*

In genetic association studies, joint modeling of related traits/phenotypes can utilize the correlation between them and thereby provide more power and uncover additional information about genetic etiology. Moreover, detecting rare genetic variants are of current scientific interest as a key to missing heritability. Logistic Bayesian LASSO (LBL) has been proposed to detect rare haplotype variants using case-control data, i.e., a single binary phenotype. LBL has been also recently extended to model two correlated binary phenotypes. As there is currently no haplotype association method that can handle a binary (B) and a continuous (C) phenotypes jointly, we extend LBL to fill this gap and refer to it as bivariate LBL-BC. We develop a bivariate model by using a latent variable to induce correlation between the two outcomes and fit the model using Markov chain Monte Carlo methods. We carry out simulations to investigate the bivariate LBL-BC and compare with the univariate LBL and bivariate LBL with two binary phenotypes. The bivariate LBL-BC performs better or similar to the other two versions. It has the highest gain in power when a haplotype is associated with both traits and it affects at least one trait in a direction opposite to the direction of the correlation between the traits. Finally we apply the method to a genome-wide association dataset on lung cancer and smoking.

Keywords: Bayesian Logistic LASSO, Genetic Association, Markov chain Monte Carlo, Lung Cancer, Smoking.

Invited Speech (IS3)

**SMALL AREA ESTIMATION OF PREVALENCE USING HIERARCHICAL AND
EMPIRICAL BAYES METHODS WITH APPLICATIONS TO UNINSURED
MINORITY SUB-POPULATION AND TOBACCO USE AMONG NURSES**

Karabi Nandy

*Department of Biostatistics and Epidemiology, University of North Texas Health Science Center,
Fort Worth, TX 76107, USA.
karabi.nandy@unthsc.edu*

What happens when national surveys fail to provide adequate information at targeted sub-population levels, such as state, county, and minority group. Few surveys are large enough to be capable of providing reliable subpopulation estimates, even if the overall sample is large. Small area estimation is a branch of survey sampling which pools information from multiple small areas with some unifying characteristics using sophisticated mixed effects models. The problem is particularly challenging for non-linear models with unit level data where conventional mixed effects models are usually inadequate. In this work a method using Hierarchical and Empirical Bayes methods was used to specifically address the small area estimation problem for unit level data for binary outcome. The proposed method can be applied to obtain state-level estimates of smoking prevalence among nurses from the Tobacco Use Supplement of the Current Population Survey (TUS-CPS). A second example, which may also have strong health policy implications, consists of estimation of the proportion of persons without health insurance for minority subpopulations. Finally, a Bayesian diagnostic procedure was implemented to check the fit of the model to data by drawing simulated values from the posterior predictive distribution and comparing these samples to observed data.

Keywords: Gibbs Sampler, Hierarchical Bayesian random-effects model, Markov Chain Monte Carlo Methods, Small Area Estimation, Tobacco.

Invited Speech (IS4)

GOODNESS OF FIT FOR A JOINT MODEL PERTAINING TO MEDICAL DATA

M.R. Sooriyarachchi* and Nayani Adikari

Department of Statistics, Faculty of Science, University of Colombo, Colombo 07, Sri Lanka.
*roshini@stat.cmb.ac.lk

In the area of Medicine, particularly Epidemiology, often responses on survival and count of patients with a specific disease are encountered. These two types of responses are usually correlated, and thus, it is more efficient to model the responses together in a joint model rather than modeling them using univariate models. The situation is more complicated when the responses are not only clustered within patients but the patients are also clustered within larger units. Lately, our group of research students has developed two models for this purpose. Of the two, we have selected the Generalized Linear Mixed Model (GLMM) with a single random effect at the larger unit level due to its computational simplicity. One discovery made by this research group was the dearth of a Goodness of Fit (GOF) test for these models. Thus, the objective of this research is to develop a goodness of fit test, examine its properties and to apply this GOF test to a real data set. The survival data are discretized and a Discrete Time Hazard Model (DTHM) is used for the first response and a Negative Binomial (NB) model is used for the count data. These two responses are used in a joint model within the framework of a GLMM model. Then, the expected values from these responses are obtained separately, and using the Hosmer-Lemeshow test the indicator variables are found and the joint model refitted with the indicator variables. The significance of these indicator variables gives a test for determining the GOF. Simulation is used to examine the properties of the test. The sample size (in terms of number of clusters and the cluster size) and the intra-cluster correlation (ICC) are varied, and the type one error and the power of the test are examined. Finally, the test is applied to a Dengue data set. The important conclusions are that out of thirty combinations twenty eight combinations, give type one error within the 95% probability interval taking the level of significance as 5%. The power is high for moderate to large sample sizes. For the data set, the joint model taking the random effect to be district and fits well as determined by the GOF. Academically, in the literature, there is no GOF test developed for this scenario. Thus, this is a novel development. Practically this development, which is often encountered in epidemiology, and can also be applied in biology, education and sociology, to name a few areas.

Keywords: Generalized Linear Mixed Model (GLMM), Goodness of Fit (GOF), Cluster, Discrete Time Hazard Model (DTHM), Dengue Epidemiology.

Invited Speech (IS5)

**JOINT MODELING OF DENGUE SURVIVAL AND COUNTS WITH JOINT
GENERALIZED ESTIMATING EQUATIONS**

J.C. Hapugoda¹ and M.R. Sooriyarachchi²

¹*Department of Organizational Studies, The Open University of Sri Lanka, Nawala,
Nugegoda 11222, Sri Lanka.*

²*Department of Statistics, Faculty of Science, University of Colombo, Colombo 07, Sri Lanka.
jchap@ou.ac.lk

Dengue is endemic in Sri Lanka as in other tropical countries. Thus, it is important to determine the factors that affect both the incidence of dengue (count) and the survival of dengue patients. A literature review carried out on this subject shows that in addition to patient characteristics, climatic variables such as rainfall, temperature and humidity, are also risk factors. In many cases, survival and incidence can be correlated in a way that, diseases that occur rarely can have shorter survival times or vice versa. Due to this fact, joint modeling of these two variables will provide interesting and certainly improved results than modeling these separately. Though joint modeling has been commonly applied in longitudinal survival data, it is not common in the literature to see survival and count variables considered together as a bivariate response in a joint model. This is often further complicated by the data being of a hierarchical nature (in the form of clusters), where there exists a geographical variation of the intensity of the disease. In such a scenario, it is better to model survival time and count data together to take account of the clustering. Thus, the objective of this study is to develop such a model in order to assess the impact of climate factors on Dengue survival and count. The method utilized in this study is Joint Generalized Estimating Equations (GEE) as it is considered as a better approach than the full likelihood approaches for modeling higher order associations in correlated data. Since the survival variable is continuous in nature with censoring information associated with it, the survival variable was converted into a discrete time approach, as it is the easiest and the most common method to analyze survival data. Thus, an approach which is based on the data structure used in Discrete Time Hazard Model (DTHM) was utilized. The restructured survival data together with counts was analyzed using response-specific marginal regression models that were estimated via GEEs, where correlation parameters were modeled through a composite correlation matrix and estimated via method of moments. The results reveal that rainfall, rainfall lag1, temperature and humidity have significant impact on Dengue survival and counts.

Financial assistance from the University of Colombo (Grant No 2015/MPHIL/S/162) is acknowledged.

Keywords: Dengue, Generalized Estimating Equations, Joint modeling.

Invited Speech (IS6)

COMPARISION OF JOINT MODELING APPROACHES FOR ANALYZING CORRELATED DISCRETE COMPETING RISK AND CONTINUOUS OUTCOME

G.H.S. Karunarathna and M.R. Sooriyarachchi

Department of Statistics, Faculty of Science, University of Colombo, Colombo 07, Sri Lanka.

**hasani@stat.cmb.ac.lk*

Clustered multiple outcomes are often encountered in many health related studies to recognize the association and to take better prevention from diseases. The common approach to those outcomes is to model each outcome separately ignoring their association among responses. Competing risk is an attractive tool in health science since it is a combination of survival time with several destination simultaneously. Thus, we proposed a novel joint modelling approach for discrete competing risk with another continuous response under the clustering nature through joint random effect model and Gaussian copula methods. Joint association in the proposed method is built through two sub models called multilevel utility model for discrete competing risk and multilevel normal model for continuous response and joining through random effects. Comparision was made through the simulation study by changing the parameter of sample size, variation of random effects and treatment effect. Simulation study reveals that copula model gives better result due to some drawbacks of random effects model. The application of dengue epidemiology in Sri Lanka proved the results of simulation study.

Keywords: Multilevel, Competing risk, Random effect, Copula .

Invited Speech (IS7)

CAN CLIMATE FACTORS INFLUENCE CKD OF SRI LANKA?

K.W.G. Rekha Nianthi^{1,*} and Dhammadika Dissanayake²

¹*Department of Geography, Faculty of Arts, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

²*Department of Pathology, Faculty of Medicine, University of Peradeniya,
Peradeniya 20400, Sri Lanka.
rekhanianthi@yahoo.com

Chronic Kidney Disease of unknown origin (CKDu) affects some agricultural communities in the equatorial region with a tropical climate. CKDu is assumed to be multifactorial and is a public health concern in North Central Region of Sri Lanka. Literature shows that the risk of developing CKDu increases with certain agricultural practices that indicate more intense physical work and it could be a form of heat-stress nephropathy arisen in part due to rapidly changing climate. This study examined the relationship between CKDu and climate variables in the last few decades. A perceptual approach to cross-sectional descriptive study was carried out on 78 road construction workers from the CKDu high prevalent area, through the pretested structured questionnaire to assess the awareness, knowledge and conventional risk factors of CKDu. The respondents were screened for CKD by a serum creatinine test. Focus group discussions were carried out to examine the contextual evidence related to CKDu. The study noticed the higher mortality due to CKDu in the months of November to January. High annual rainfall was observed with a lower incidence of kidney disease in the subsequent year. The maximum average temperature is ranging between 31.2 - 37.3 °C during March and April. The highest number of patients was observed when the annual average temperature exceeds 27.6 °C. The correlation between temperature and mortality showed a positive relationship with weak associations. A significant increasing trend (p -value < 0.01) and a decreasing trend (p -value < 0.05) were identified during January and May, respectively. The wind speed indicates a significant decreasing trend. Annual THI, is higher in the more recent period of 2009-2016 with ($THI > 26$) and therefore study area is recognized as “less comfortable zone” for human heat stress. All the respondents were males (78) and 6% engaged in indoor work, 85% outdoor work, and 9% engaged in both indoor & out door. When tested with serum creatinine, 21% respondents had increased serum creatinine indicating presence of CKD, which is higher than age and sex specific prevalence rates in Medawachchiya (15%). Out of heat stress symptoms, headache was the commonest (50%) and muscle cramps (32%), urine burn (27%), fainting (20.5%), dizziness (10.2%) and vomiting (3.8%) were the other important symptoms present. Symptoms were more common from April to August. None of the participants were aware of heat stress and its impact. However, most of them get rest in the shade after working for about 2 - 3 hours and were not aware of preventive measures for heat stress. The study indicates the importance of screening the outdoor workers for CKD. and the finding will be very useful for planning of future mitigations.

Keywords: CKD, Climate, Rainfall, Temperature, THI.

Invited Speech (IS8)

NON-METRIC TOOTH CROWN TRAITS IN A CONTEMPORARY AND TWO ABORIGINAL POPULATIONS IN SRI LANKA: COMPARISON WITH OTHER WORLD POPULATIONS

H.R.D. Peiris^{*}, J.A.A. Jayaweera, A.K.S. Arambawatta and C.D. Nanayakkara

*Department of Basic Sciences, Faculty of Dental Sciences, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

^{*}rdpeiris@gmail.com

Comparison of nonmetric dental features has been valuable in making assessments of the affinities of human populations. This study was conducted to determine the frequencies of non-metric tooth crown traits in a contemporary and two aboriginal Vedda populations (Dambana and Anuradhapura) in Sri Lanka and to investigate the affinities of these morphological variations with those of other world populations. Fifty dental plaster casts from each population were observed. The Arizona State University dental anthropology system was adopted for classification of the 13 traits observed. Affinities among the contemporary and two aboriginal populations of Sri Lanka and other world populations were expressed in two dimensions of the principle coordinate analysis. Cusp number in mandibular second molar and Hypocone absence in maxillary second molar had the highest frequency and Shoveling and Double Shoveling in the maxillary central incisor had the lowest frequency in all three populations. Contemporary Sri Lankans had higher occurrence of Cusp of Carabelli's in maxillary first molar and Deflecting wrinkle in mandibular first molar. Dambana Vedda had higher prevalence of Cusp 7 in mandibular first molar. Anuradhapura Vedda had lower prevalence of Y groove pattern in mandibular second molar, Interruption groove in maxillary lateral incisor and Cusp 6 in mandibular first molar. The principal coordinate analysis showed that the contemporary Sri Lankans, Dambana and Anuradhapura Vedda located with the Western Eurasia population groups. While showing close affinities with early South Asian populations, Dambana and Anuradhapura Veddas were deviating away from the contemporary Sri Lankans in the second principal coordinate axis showing their genetic drift from the modern Sri Lankans.

Keywords: Tooth morphology, Non metric, Sri Lankan Sinhalese, Aboriginal Vedda.

ENVIRONMENTAL STATISTICS

Abstract No: 01

**IMPACT OF ENERGY CONSUMPTION AND ECONOMIC GROWTH ON
CARBON DIOXIDE EMISSIONS IN SRI LANKA**

K.A.U. Sakunthala* and T.S.G. Peiris

*Department of Mathematics, Faculty of Engineering, University of Moratuwa,
Moratuwa 10400, Sri Lanka.
sakunthalaudari@gmail.com

Globally proliferation of climate change and global warming has been deliberated as a controversial environmental issue in sustainable economic growth in recent decades. Upsurge in economic activities and global trades have caused significant increase in CO₂ emissions triggering deteriorating and irreparable ecological damages, affecting ecosystem functions, and viability of species and human well-being, since almost 80% of the world's energy demands is contented by fossil fuels. The studies of identifying 'energy-economic-CO₂ emissions nexus' have been given high priority by many countries, and facilitated in mitigation targets of continuous reduction in CO₂ emissions and sustainable policy implementations. However, studies on this aspect have not much been carried out in Sri Lanka. This study is therefore utilized the Vector Error Correction Model (VECM) framework to investigate "impact of energy consumption and economic development on CO₂ emissions in Sri Lanka" using annual time series data over the period of 1971 – 2014. The dataset was obtained from the World bank data base. The Johansen cointegration tests suggested an existence of unique cointegration vector for long-run cointegration relationships among tripartite, and the fitted VECM (1) model was highly stable encompassing white noise residuals. The long-run elasticity estimates suggested significantly that per capita economic growth has a long term positive impact on per capita CO₂ emissions, which implies that economic developments in Sri Lanka leads to greater CO₂ emissions in long run effectiveness. Surprisingly, per capita fossil fuel energy consumption influenced negatively on per capita CO₂ emissions in long run association. The reason may be that the fossil fuel energy consumption in Sri Lanka is well below to trigger any significant CO₂ emissions compared to developed nations. It is recommended that similar studies to be performed in regular intervals.

Keywords: CO₂ emissions, Cointegration, Economic Growth, Energy Consumption, VECM.

Abstract No: 04

**UNDERSTANDING CHRONIC KIDNEY DISEASE IN HISTORICALLY
GROUNDWATER RELIANT COMMUNITIES IN SRI LANKA: A
SOCIOECONOMIC APPROACH**

K. Kafle^{1,*}, S. Balasubramanya¹, T. Horbulyk^{1,2}, D. Stifel^{1,3} and H. Manthrithilake¹

¹*International Water Management Institute(IWMI), Battaramulla 10120, Sri Lanka.*

²*University of Calgary, Alberta, T2N 1N4, Canada.*

³*Lafayette College, Easton, PA 18042, USA.*

^{*}*k.kafle@cgiar.org*

Chronic Kidney Disease (CKD), a serious health problem in Sri Lanka, has affected large sections of the rural population and has placed a burden on scarce public health resources. CKD of unknown etiology (CKDu) is a distinctive form of CKD that has grown in importance over time and whose causes are not well understood. It is hypothesized that CKDu is caused, in part, by consumption of groundwater. However, the hypothesized connection between groundwater use and CKDu in Sri Lanka has not, until now, been well supported by data that connect CKD prevalence with historical and sustained use of groundwater sources. This analysis provides new estimates of Chronic Kidney Disease (CKD) prevalence, and examines the relationship between CKD and rural households' historical reliance on groundwater consumption and historical household behaviours across ten districts most affected by CKD in Sri Lanka. A carefully designed survey of 8,000 households provides information on whether these households self-reported having a member who had been clinically diagnosed with CKD. Households were classified according to whether or not they had used groundwater as their primary source for drinking or cooking for at least five years between 1999 and 2018. In-depth personal interviews were conducted with members of 1,497 households to elicit household behaviours related to historical choices of drinking water source, water treatment practices and agrochemical use. Data were analysed using probability weighted means, two sample t-tests, and probit regressions. Results show that more than 98% of both CKD affected and unaffected households had consumed groundwater as the primary source for drinking or cooking. CKD prevalence – the share of affected households in the most affected areas – was 15.4% and a CKD affected household, on average, had more than one symptomatic adult (~1.25). Using recall data for the last 18 years, the results illustrate how water source choices have changed over time for rural households. Over time, the share of households choosing a publicly provided reverse osmosis unit as its primary water source has increased, while the share of households using untreated groundwater has fallen. The share of households using public sources of drinking water has also increased due to the move from private household wells to publicly provided reverse osmosis units. When comparing CKD and non-CKD households, no major differences were observed in their historical choices of households' primary water sources by type, duration of use and timing of source-type by year.

Financial assistance from the United States Department of State (Grant No. SLMAQM17GR2021) is acknowledged.

Keywords: Chronic Kidney Disease, Groundwater Reliance, Historical Behaviours.

Abstract No: 05

**GROUNDWATER QUALITY AND ALTERNATIVE SOURCES OF
GROUNDWATER IN MOST CHRONIC KIDNEY DISEASE AFFECTED
COMMUNITIES IN SRI LANKA**

S. Balasubramanya^{1,*}, T. Horbulyk^{1,2}, K. Kafle^{1*}, D. Stifel^{1,3} and H. Manthrithilake¹

¹*International Water Management Institute(IWMI), Battaramulla 10120, Sri Lanka.*

²*University of Calgary, Alberta, T2N 1N4, Canada.*

³*Lafayette College, Easton, PA 18042, USA.*

^{*}*s.balasubramanya@cgiar.org*

Chronic Kidney Disease (CKD) has affected large sections of the rural population in Sri Lanka and has placed a burden on scarce public health resources. CKD of unknown etiology (CKDu) is a distinctive form of CKD that has resulted in considerable morbidity and mortality among the community and whose causes are not well understood. Though it is assumed that CKDu is caused, in part, by consumption of groundwater, the connection between groundwater use and CKDu in Sri Lanka has not, until now, been well supported by data that connect historical and sustained use of groundwater sources with meaningful measures of disease prevalence at the household level. This analysis draws from a carefully designed survey of 1500 households (both CKD affected and non-affected households) and water testing results from 240 household wells that had been primarily used by both CKD affected households (CKD wells) and non-affected households (non-CKD wells). The analysis first examines systematic differences in water quality between CKD and non-CKD wells and then presents current snapshot of alternative sources of groundwater being used by both CKD and non-CKD households in the most CKD affected areas. The survey also elicits householder's preference for alternative sources of groundwater and their trust in the providers of drinking water. Data were analysed using probability weighted means, two sample t-tests, and probit regressions. Results show that all elements tested for in each of the 240 wells meet the Sri Lankan Standard for potable water. Based on a single test of each well in 2019, there is no systematic difference in water quality between CKD wells and non-CKD wells. Among the five alternative sources of groundwater provided – Reverse Osmosis (RO), rain water harvest, municipal water, bowser/tanker water and public water supply, RO is the most popular choice. About 49% of households used RO as their primary source of drinking water in 2018. On average, they spend around 17 minutes per trip to fetch water from RO plants. About 28% of households do not pay for RO water, and for those without free access, the average price of RO water is LKR 1.11/liter. In addition, RO water is perceived as the safest alternative source of drinking water by households previously reliant on groundwater as their primary source. Nonetheless, the well testing results should be taken with caution because the water samples were drawn from the bottom of the wells at a single point in time which cannot capture water quality variation across seasons and different well depths.

Financial assistance from the United States Department of State (Grant No. SLMAQM17GR2021) is acknowledged.

Keywords: Chronic Kidney Disease (CKD), Groundwater quality, Alternative sources.

Abstract No: 07

**FITTING A GENERALIZED LINEAR MODEL FOR RAINFALL OCCURRENCE
AND RAINFALL AMOUNT IN COLOMBO DISTRICT**

A. Selvaratnam* and P. Wijekoon

*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

*ahthigas@gmail.com

Rainfall modelling possesses a major role in predicting and forecasting rainfall to be utilized in many sectors such as agriculture and hydrology. The variables that are considered for rainfall modelling in most of the studies are rainfall occurrence, rainfall amount and rainfall intensity. Since the rainfall occurrence is a discrete variable, the Poisson distribution and the method of Markov chain are the widely used methods for modelling. The Gamma distribution, Exponential distribution and Mixed Exponential distribution are the common distributions applied to model the rainfall amount and rainfall intensity. In this study, an attempt has been made to fit a single distribution to model both rainfall occurrence and rainfall amount simultaneously, by using Generalized linear models with family of Tweedie distribution. Daily rainfall amounts in Colombo district from 1951 to 2000 were used for model fitting and the relevant data from 2001 to 2007 were used for model validation. To study the cyclical patterns and seasonal variations of daily rainfall amount, five models with different sets of predictors were fitted, namely annual frequency sine-cosine terms, six-monthly frequency sine-cosine terms, monsoon factor, month factor and six-monthly frequency sine-cosine terms with month factor. The model with sine and cosine terms for six-monthly frequency fitted well to the data based on the mean daily rainfall amounts for 50 years. The best fitted model produced continuous values indicating the rainfall events for which exact zero values indicate the non occurrence of rainfall. Simulated rainfall data for the period from 2001 to 2007 resembled the actual data in terms of summary statistics. This model can be further improved by adding more predictors such as climate indices.

Keywords: Gamma distribution, Generalized linear models, Poisson distribution, Tweedie distribution, Zero-inflated.

Abstract No: 08

PREDICTING ANNUAL STEM DIAMETER INCREMENT OF SELECTED TREE SPECIES IN SINHARAJA RAIN FOREST CONSIDERING TREE AND STAND LEVEL EFFECTS

D.P.V.R. Dissanayake* and P. Wijekoon

*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.
vishmidissanayake95@gmail.com

The information on the diameter increment of tree species is important for developing a sustainable forest harvesting plan and for reforestation. Fitting a forest growth model for diameter increment can be utilized to get this information. The objective of this study was to develop a predictive model for the annual diameter increment of a multispecies stand in the Sinharaja Rain Forest in Sri Lanka. In order to fulfill this objective, four linear mixed-effect models were fitted to predict annual diameter at breast height (dbh) increment of trees by using dbh, the natural logarithm of dbh, and species as a random effect. The model that contained random coefficients for the intercept term, dbh and log (dbh) was selected as the best-fitted model based on the Akaike Information Criterion (AIC) values and the likelihood ratio test. Then, the best-fitted model was expanded by adding other subsequent variables that describe the effects of competition from surrounding trees and a size structure component, which is the maximum dbh. The final model comprised of dbh, log(dbh), stocking density, maximum dbh and competition from trees smaller than the subject tree. Using the best fitted final model, the annual dbh increment of the 10 selected abundant species was calculated. It was noted that all the selected species have growth rates that increased rapidly at the initial stage, and later reach its unique maximum growth rate. These increment patterns reflected that different species followed different annual dbh increment patterns. Therefore, this final model can be consolidated into an effective empirical model to project the future growth of a tropical rain forest.

Keywords: Diameter increment, Mixed-effect model, Diameter at breast height, Empirical model, Random effects.

Abstract No: 12

THERMAL COMFORT AND CLIMATE REGIONALIZATION IN SRI LANKA

H.M.N.S. Kumari^{*} and P. Wijekoon

*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.
nadeeshabandara20@gmail.com

Human and climate are related such that human response to the climate changes by changing their lifestyle. Therefore, it is important to have knowledge about climate and comfort regions as this directly affects day-to-day life and for the tourism industry. The Effective Temperature (ET), Temperature Humidity Index (THI) and Relative Strain Index (RSI) are important indices when identifying thermal comfort regions. However, in previous studies, only the THI has been used for this purpose. Therefore, in this study, the association of ET, THI and RSI with thermal comfort was examined. The three indices were computed from the monthly air temperature and relative humidity between 2009 and 2017 for 30 meteorological and agrometeorological stations in Sri Lanka. Three scatter plots were drawn by considering the elevation of each station versus the mean of the respective index. According to these scatter plots, it was noted that the ET, THI and RSI values decrease as the elevation increases. A cluster analysis was used to classify the stations to form six thermal climatic regions. The complete linkage, average linkage, centroid and Ward's methods all produced the similar 6 cluster solutions with the highest Dunn index of 0.32. Each of these clusters represented a thermal comfort region and named them as 100% cold stressed & discomfort, slightly comfort, comfort, extremely comfort and 100% unstressed, moderately comfort, and 100% heat stressed and discomfort. Among the 30 meteorological and agrometeorological stations, two are found as 100% cold stressed and discomfort, three as slightly comfort, one as comfort, three stations as extremely comfort and 100% unstressed, five stations as moderately comfort, and the remaining 16 are 100% heat stressed and discomfort. Most of the stations in the wet zone and dry zone are observed as discomfort, and ET, THI and RSI values were also higher for the weather stations in these zones. Further, a monthly classification was done to identify the thermal comfort of each month relevant to the thermal comfort regions. It was identified that January is the most moderately comfort month while the months from May to August are 100% heat stressed and discomfort months.

Keywords: Climate Regionalization, Cluster Analysis, Comfort Regions, Thermal Comfort.

Abstract No: 13

DEVELOPMENT OF LONG MEMORY MODEL TO FORECAST WEEKLY RAINFALL

H.P.T.N. Silva* and T.S.G. Peiris

¹*Department of Social Statistics, Faculty of Humanities and Social Sciences,
University of Sri Jayewardenapura, Nugegoda 10250, Sri Lanka.*

²*Department of Mathematics, Faculty of Engineering, University of Moratuwa,
Moratuwa 10400, Sri Lanka.*

^{*}*thanuja@sjp.ac.lk*

Awareness of pattern of weekly rainfall and its variability facilitate to make effective decisions with respect to climate monitoring. Though various statistical and non statistical techniques have been developed for rainfall modeling with increasing degree of accuracy, there is still a noticeable gap for prediction of rainfall. The aim of this study was to model weekly rainfall in context of long memory along with the conditional heteroskedasticity. Weekly rainfall data (1990-2017) in Colombo city was obtained from the Department of Meteorology, Sri Lanka. Of the various types of long memory models developed for weekly series, the best fitted model is ARFIMA-GARCH for deseasonalized data. The model was trained using weekly rainfall data from 1990 to 2014 and validated using weekly data from 2015 to 2017. The forecasting performance of the new model is not much diluted with the increase of the forecasting length. The exact maximum likelihood estimation method was utilized to estimate the model parameters, and Monte Carlo simulation was carried out with various fractional differencing parameters to evaluate the suitability of the estimation method. The simulation study provided the empirical evidence to optimal accuracy of parameter estimation. The best fitted model developed is ARFIMA-GARCH for deseasonalized data. The forecasting performance of the model was evaluated based on the novel index developed using absolute error for an independent data set in addition to the classical indicators. The novel long range dependency model is recommended to be used in forecasting weekly rainfall in Colombo city in Sri Lanka.

*Financial assistance from the University of Sri Jayewardenepura
(Grant No ASP/01/RE/HSS/2016/75) is acknowledged.*

Keywords: ARFIMA-GARCH, Forecasting, Fractional differencing, Long-memory,
Weekly rainfall.

Abstract No: 14

PUBLIC ACCEPTABILITY OF TREATED SURFACE WATER FOR POTABLE PURPOSE: CASE STUDY IN VAVUNIYA

M.A.O.D. Mathara Arachchi* and S. Devaisy

*Department of Bio-Science, Faculty of Applied Science,
Vavuniya Campus of the University of Jaffna, Vavuniya 43000, Sri Lanka
okiladulanjali@gmail.com

The low public acceptance is one of the major challenges in implementing and succeeding surface water intake projects for potable purpose. The water treated from Per-Aru reservoir is the source of water supplied by the National Water Supply and Drainage Board (NWSDB) for some parts of Vavuniya district. The low public demand towards the use of treated water is one of the obstacles faced by the NWSDB in operating the plant. Thus, the objective of this study was to assess the public acceptability towards the use of Per-Aru treated water for potable purposes in Marakarampalai area located in North-west part of Vavuniya. This study area was selected as it purely receives treated water from Per-Aru Water Treatment Plant for water supply. Questionnaire survey was conducted between two groups of sample population; beneficiaries (those who have NWSDB water connection, n = 40), and non-beneficiaries (those who do not have NWSDB water connection, n = 40) in Marakarampalai area. The questions were focused to assess the satisfaction towards the treated water among the beneficiaries and to assess the willingness/constraints to obtain water supply among non-beneficiaries. Results showed that half of the beneficiaries use the treated water for potable purposes as they are aware of the periodic water quality checkups. However, the rest only uses for non-potable purposes as they do not trust the quality of water for potable purposes. More than half of the beneficiaries obtained the water supply connection mainly to get rid of water shortage issues in dry season. It was noticed that the people those who have access to alternative water sources are not willing to get water connection paying extra bills. For non-beneficiaries, financial constraints and lack of awareness are the major factors to get the water connection. The provision of financial benefits and environmental education programmes by the NWSDB can be encouraged to enhance the public demands towards the use of treated water.

Keywords: Acceptane, Per-Aru, Potable, Public, Treated water.

Abstract No: 16

ANALYSIS OF ASSOCIATION BETWEEN DENGUE CASES AND RAINFALL IN SRI LANKA

S.M.J.H. Nisansala* and P. Wijekoon

*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.
jaminthahashini@gmail.com

Dengue fever is the most significant mosquito-borne disease caused by a virus. *Aedes* mosquitoes are the vectors of the dengue virus. The lifestyle of mosquito is influenced by climatic factors such as rainfall, temperature and humidity. The main objectives of this study were to analyse the relationship between dengue cases and climatic factors using circular statistics and to find empirical models for estimating dengue cases using climatic factors. Monthly dengue incidents were collected from the website of the Epidemiology Unit in Sri Lanka, and rainfall, temperature, and humidity data from January 2010 to December 2017 were obtained from the Department of Meteorology, Sri Lanka. Comparison of dengue cases and rainfall amounts using rose diagram reveals that months with higher rainfall do not show higher number of dengue cases. However, after time lag of 1, 2, or 3 months of rain, there is an increment in dengue cases. By considering these circumstances, empirical models were developed using 0, 1, 2 and 3 time lags for each district with the aid of generalized linear negative binomial model. For each district, four models based on time lags were fitted separately including rainfall, temperature and humidity as predictor variables, and another four models by adding monsoon season as an additional categorical covariate to understand the seasonal effects. The best-fitted distribution for each case was identified by comparing the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). When considering models without monsoon seasons as a covariate, lag 3 model is a perfect match for Ratnapura and Colombo districts; Lag 2 model is a perfect match for Anuradhapura, Badulla, Kurunegala and Kandy districts and lag 0 model is a perfect match for Gampaha district. When considering models with monsoon seasons; Lag 3 model is a perfect match for Ratnapura district; lag 2 model is a perfect match for Anuradhapura, Badulla and Kandy districts; Lag 1 model is a perfect match for Kurunegala district; Lag 0 model is a perfect match for Colombo and Gampaha districts. According to the results, only Kurunagala and Colombo districts showed different results when adding monsoon season as a covariate. Finally, the validity of the fitted models was compared with actual data.

Keywords: Circular statistics, Climatic factors, Dengue cases, Generalized linear negative binomial model, Rose diagram.

Abstract No: 18

DATA EXPLORATION AND PRE-PROCESSING TECHNIQUES ON AIR POLLUTION AND METEOROLOGICAL DATA IN SYDNEY REGION

W.M.L.K.N. Wijesekara^{*} and L. Liyanage

*School of Computing, Engineering and Mathematics, Western Sydney University,
Locked Bag 1797, Penrith NSW 2751, Australia.*

^{*}18570263@student.westernsydney.edu.au

Data preparation typically consumes 80-90% of the total time taken to complete a data mining project. It is a crucial step as the performance of any model highly depends upon the quality of the data. “Garbage in, Garbage out (GIGO)”. One of the major issues in data pre-processing stage in a large dataset is missing values. Air pollution and meteorological data typically consist of many missing values. Proper imputations should be carried out to avoid any bias caused by missing values. The main objective of this study was to propose suitable techniques to be used in data preprocessing for air pollution and meteorological data in Sydney region, Australia. The dataset consists of hourly measurements of air pollution and meteorological variables from 1994-01-01 01:00:00 AEST (Australian Eastern Standard Time) to 2018-12-31 24:00:00 AEST recorded at each station in Sydney Region. The preprocessed data can be used in spatiotemporal analysis to assess the impact of climate change on different health aspects. Principal Component Analysis (PCA) was used to analyze the relationships of variables. Highly positively-correlated variable groups were [CO, NO, NO₂], [O₃, temperature, wind speed], [Visibility, PM2.5, PM10] and [wind direction, humidity]. Humidity was highly negatively correlated with O₃ and temperature. Further, 82% of the total variation is explained by the first five principal components. Six well-established techniques to impute missing values in time series data; Mean Imputation, Spline Interpolation, Simple Moving Average, Exponentially Weighted Moving Average, Kalman Smoothing on Structural Time Series Models and Kalman Smoothing on Autoregressive Integrated Moving Average (ARIMA) models were compared. Imputation method based on Kalman Smoothing on Structural Time Series model showed better performance over the other methods for missing values under Missing Completely at Random (MCAR) mechanism for the data obtained in Sydney area.

Keywords: Imputation, Missing Completely at Random, Smoothing, Spatiotemporal Analysis.

Abstract No: 19

**WASTE-DERIVED BIOSORBENTS TO REMOVE ORGANICS FROM
RICE MILL WASTEWATER**

U.D.T.M. Uduwarage* and S. Devaisy

Department of Bio-science, Faculty of Applied Science, Vavuniya Campus of the University of Jaffna, Vavuniya 43000, Sri Lanka.
*thamali.uduwarage12@gmail.com

The objective of this study was to compare the performance of natural biosorbents (rice hull and peanut hull) in removing organic pollutants from rice mill wastewater. The biosorbents were prepared through the processes of washing, drying, grinding and sieving to obtain particle size of less than 1 mm. Batch adsorption experiments were conducted for rice hull and peanut hull separately for varying doses (0 - 6 g/L) and contact times (0 - 300 min) under similar conditions to determine the optimum dosage and optimum contact time at pH 8.0 ± 0.25 and at temperature 28 ± 2 °C, respectively. The dose of 0 g/L was considered as the control of the experiments. The performance efficiencies of the biosorbents were tested based on the reduction of Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) from wastewater using three replicates for each test. Results revealed that the optimum doses of rice hull and peanut hull were found to be 3 g/L and 4 g/L respectively in reducing organics. At the optimum doses of rice hull and peanut hull, BOD reductions were observed to be 73% and 65%, whilst COD reductions were 73% and 70%, respectively, at a contact time of 4.0 hours. Considering the optimum contact time at the respective doses, the rice hull achieved optimum removal at 120 mins, whilst the peanut hull achieved at 180 mins at the shaking of 150 rpm. Based on a two-way t-test, the rice hull and the peanut hull significantly remove organics from the ricemill wastewater (p -value < 0.05). The findings showed that the rice hull is more efficient in the removal of organics than peanut hull as rice hull requires lesser dose and contact time. Furthermore, it was noticed that the rice hull is capable of adsorbing organics to its binding sites are faster compared to the peanut hull. Therefore, considering the cost-effectiveness and promoting waste to resource concept, the waste-derived biosorbents, such as rice hull and peanut hull, can be used as potential biosorbents to remove organics from ricemill wastewater.

Keywords: Biosorbents, Organics, Peanut hull, Rice hull, Wastewater.

Abstract No: 24

SHORT TERM ELECTRICITY DEMAND FORECASTING FOR GENERATION SCHEDULING: A CASE STUDY IN SRI LANKA

E.D.T. Somarathne^{1,*}, D.S.K. Karunasinghe² and J.V. Wijayakulasooriya³

¹*Faculty of Agriculture, Rajarata University of Sri Lanka, Anuradhapura 50000, Sri Lanka.*

²*Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

³*Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

^{*}*deshani_s@yahoo.com*

Short-term electricity demand forecasting with higher accuracy is a necessity for accurate power system planning and operation to reduce costs and increase effectiveness. In Sri Lanka, one day ahead electricity demand is currently predicted based on the comparison with demand history. Therefore, this study attempted to investigate a better method for short term load demand forecasting in Sri Lanka. The objective of this study is to develop a system of seasonal Autoregressive Integrated Moving Average (ARIMA) models to forecast half-hourly electricity demand for the next 24 hours, and to evaluate its effectiveness compared to that of the currently used method. Electricity demand (in MW) recorded at half-hourly intervals in Sri Lanka from January 29, 2006 to January 28, 2007 was used for this study. The demand of first 256 days (~70% of the dataset) was used to train the methodology and that of last 109 days (~30% of the dataset) was used for the testing. Forty eight half-hourly demand series were individually analyzed to investigate within-day and within-week periodic patterns. Then, Forty Eight Seasonal ARIMA models were developed. Two currently used techniques were also implemented to find the electricity demand for the next 24 hours. Using the two types of methods, electricity demand was predicted for the week of 18th-24th December 2006 which is in the testing time period. The forecasts obtained by the two methods were compared with the actual values by using several relative and absolute error measures, namely, Normalized Root Mean Square Error (NRMSE), Coefficient of Determination, Mean Average Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Two time periods in a day (3:30 AM to 6:00 PM and 6:00 PM to 3:30 AM) were found to be represented by two different sets of ARIMA models, indicating similar dynamics in each half an hour in a given period. The existing methods were, however, not capable of identifying such patterns. The computed error measures were 0.1629, 45.17, 3.73, 37.14 and 0.9735 with seasonal ARIMA models, 0.3348, 111.07, 6.47, 69.48 and 0.8395 with existing techniques for NRMSE, Coefficient of Determination, MAPE, RMSE and MAE, respectively, showing better prediction accuracy with ARIMA models. Existing methods involving manual interpretations have difficulties in applying to longer test periods. Therefore, mimicking the existing methods closely with an automated methods and evaluating the two methods with longer training and test data sets are ongoing.

Keywords: Forecasting, Half-hourly Electricity Demand, Seasonal ARIMA.

Abstract No: 25

SPATIAL AND TEMPORAL VARIATION OF LAND SURFACE TEMPERATURE IN COLOMBO REGION

R.P.G. Rajapaksha^{1,*}, K.W.G. Rekha Nianthi² and S.I.S. Subasinghe²

¹*Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Department of Geography, Faculty of Arts, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

** gayanirajapaksharp@gmail.com*

Land Surface Temperature (LST) is the radiative skin temperature of the land surface. Urbanization has a direct adverse effect on increasing the LST of a city. Removal of rural land-cover being replaced with the non-evaporating surface has caused to reduce the evapotranspiration and increased storage and transfer of sensible heat. Therefore, urban areas have a high rate of surface air temperature relative to neighboring vegetated areas. Satellite Remote Sensing and GIS are powerful tools to identify the LST variations. Sri Lanka also experiences a messy urbanization pattern, especially, in Colombo district associated with a large number of surface covers. This urban sprawl led to increasing the LST in the overall area of Colombo district. The purpose of this research was to identify the spatial and temporal variation of LST in Colombo district from 1997 to 2017. LST values were derived by using Landsat TM 5 and 8 (OLI/TIRS) images in 1997, 2008 and 2017. The Environmental Criticality Index (ECI) was calculated and presented the values as ECI maps. Environment Critical area can be used for identifying and monitoring the areas where vulnerable to spread the heat-related epidemics, heat stress, vector-borne diseases and LST, which can be applied for the future urban planning policies in Colombo. The results revealed that LST has gradually increased in Colombo city and other suburban areas, such as Dehiwala, Moratuwa, Awissawella, and Maharagama in 2017 as compared to 1997. The mean LST in 1997, 2008 and 2017 are 26 °C, 28 °C, and 27 °C, respectively. According to ECI in 1997, 0.0048% of the area had high ECI value and 1.21% area had moderate, in 2008, 0.4% area had very high, 2.3% area had high and 13.5% area had moderate ECI values and in 2017, 0.1% area very high ECI values, 1.5% area had high and 12.9% area had moderate ECI values. According to the overall ECI results, the environmentally critical areas were identified in the Central Business District, along with the coastal belt and the main transportation network. Newly emerged high environmental critical areas were observed in 2017 in Awissawella suburban and the Seethawaka Export Processing Zone.

Keywords: Environmentally Criticality Index, Land Surface Temperature, Sensible heat, Spatiotemporal, Urbanization.

Abstract No: 27

**DETERMINATION OF PHOTOSYNTHETIC BIOMASS, OXYGEN
PRODUCTION, AND WATER PRODUCTION IN THE EARLY GROWTH
STAGES OF AGARWOOD**

H.N. Chathuranga^{1,*}, F.R. Senanayake¹, M.K.L. Chandana² and D. De Zoysa¹

¹*Earth Restoration (Pvt) Ltd, Belipola Arboretum, Mirahawatte 90134, Sri Lanka.*

²*Department of Export Agriculture, Faculty of Agriculture Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, 70140, Sri Lanka.*

^{*}nchedigalla@gmail.com

The main objectives of this study were to establish the photosynthetic biomass (PB) gains in the early growth phase of Agarwood. The study also examined pre-existing data with relations to age, tree height, crown height and tree diameter. The study was based on field data collection and through observations based on physical measurements which focused on PB accumulation in the early growth stages of Agarwood. The sampled individuals were found in the home gardens in Batuwangala, Neluwa, Sri Lanka. Sampling was carried out in four sites which represented the growth at 1, 2, 3 and 4 years. Purposive sampling was carried out and 10 individuals from each age group were selected for a total of 40 individuals of the species. The crown volume of the PB component was recorded and leaf samples were collected. Total tree height, crown height, and diameter of the stem were also measured from selected individuals. Fresh weight was measured from the leaves collected and analyzed to derive PB curves. The photosynthetic biomass value was measured based on the oxygen production (1 g of PB gives 0.4 g of oxygen) and water production (1 g of PB gives 100 g of Water). MS Excel was used to obtain graphical illustrations. PB, oxygen production, and water production of Agarwood were showed exponential growth after the age of 3 years. A regression analysis was conducted and it is found that the PB, oxygen production, and water production had a significant associations with the age. According to the PB results obtained for the Agarwood, the tree height, crown height, and tree diameter also showed an increasing trend.

Keywords: Crown height, Photosynthetic biomass, Tree diameter, Tree height

Abstract No: 33

BANANA PSEUDO STEM BIOMASS AS A PROSPECTIVE ADSORBENT TO REMOVE INDUSTRIAL DYES

B.K.Pathiraja and U.K.Jayasundara*

College of Chemical Sciences, Institute of Chemistry Ceylon, Rajagiriya 10107, Sri Lanka.

**udayaj@ichemc.edu.lk*

Most of the industrial dyes are coloured as they are used to impart various colours in industrial products. Since these dyes are harmful, toxic, and hazardous to human, it is important to eliminate such contaminants from wastewater prior to the disposal. Adsorption technique is commonly used to remove such contaminants from polluted water. In this study, the usage of banana pseudo stem was tested as a low-cost adsorbent. Banana pseudo stem was cut into pieces and sun dried for seven days. Then it was powdered, washed with deionised water and dried for 24 hours at 60 °C to prepare banana pseudo stem biomass (BSSB). The prepared BSSB was used as the adsorbent, and several industrial dyes, such as methylene blue (MB), rhodamin B (RHB), acid red 88 (AR88), and direct blue 86 (DB86) were investigated as adsorbates. The batch sorption studies were conducted in triplicate by controlling pH and contact time. The results revealed that both DB86 and RHB had the maximum adsorption capacities of (0.96±0.02) mg/g and (6.26±0.01) mg/g, respectively, at pH 2.00 while AR88 had (1.66±0.01) mg/g at pH 4.00. MB had showed the highest adsorption capacity (264.40±1.23) mg/g at pH 6.00. The equilibration time for MB was found to be 60 minutes, while 120 minutes for RHB and AR88. The kinetics study demonstrated that data were well fitted to pseudo second order method kinetics. Further, a comparison study was performed to compare the adsorption capacities of BPSB with that of activated charcoal (AC). It showed that adsorption capacity of MB on BPSB and AC was almost the same. However, the adsorption capacity of AR88 on BPSB was remarkably higher than that of AC. Further, adsorption capacity of DB86 on BPSB was almost 10 times higher than AC. In addition, adsorption capacity of RHB on BPSB was 4 times that of AC according to the results obtained. According to the results, it can be concluded that BPSB would be a promising candidate for industrial dye adsorptions.

Financial assistance from the Institute of Chemistry Ceylon (Grant No R-18-09) is acknowledged.

Keywords: Adsorption, Biomass, Industrial dyes, Pseudo Second Order kinetics, Waste water.

Abstract No: 39

A STUDY OF VARIATIONS IN THE RAINFALL OF THE WESTERN PROVINCE BASED ON REPRESENTATIVE CONCENTRATION PATHWAY SCENARIOS

L.K.C. Tillekeratne^{1,*}, W.C.W. Navaratna² and G.D.P. Dharmarathna³

¹*Sri Lanka Technological Campus, Padukka 10500, Sri Lanka.*

²*The Open University of Sri Lanka, Nawala, Nugegoda 10250. Sri Lanka.*

³*Department of Meteorology, Bauddhaloka Mawatha, Colombo 07, Sri Lanka.*

^{*}*kumtarzi@gmail.com*

Western Province, Sri Lanka is a densely populated area that may face future extreme weather events due to variations and extremities in monsoonal rainfall. This study explores variations of the predicted rainfall in the Western Province, Sri Lanka, from 2016 to 2050, based on Representative Concentration Pathway (RCP) projections proposed by the Intergovernmental Panel on Climate Change, in the Fifth Assessment Report. The monthly rainfall data series observed from 1986 to 2015 pertaining to five major cities Colombo, Kalutara, Katunayake, Labugama and Rathmalana were used as historical data. Model selection was done by comparing historical data with data generated from six models (CANE, MRI, GDFL, CNRM, NCAR and CSI from the NEXX GDDP data series of the Coupled Model Inter-comparison Project Phase 5) for RCP values 4.5 and 8.5. Mean Average Deviation, Mean Absolute Percentage Error, Root Mean Squared Error and visual analysis were used as evaluation criteria for model selection. The models selected were used to calculate the 90th percentile and 10th percentile for rainfall extremes, monthly averages, anomalies and annual rainfall volume. An important finding from the study is that urbanized cities show positive trends in heavier rainfall except for Labugama. Other remarkable observations were that, while the 1st monsoon and 2nd inter monsoon gradually accumulate around July and October, and the total annual rainfall volumes in the urban regions decrease, Labugama, a vegetation rich region, shows an increase. These findings lead to the hypothesis that, vegetation induced micro climate's influence, plays an important role in climate change, apart from the variations due to Greenhouse gas emission trapped radiative forcing captured by the RCP scenario models. These observations need to be further studied since improper land and utility usage of the province may have a detrimental effect on the life of its' future population as well.

Keywords: Rainfall Variations, Rainfall Extremes, Precipitation, Representative Concentration Pathways.

Abstract No: 40

USING BLACK-BOX MACHINE LEARNING TECHNIQUES TO IDENTIFY SPATIAL DEPENDENCE IN OCCUPANCY DATA

N.M. Mohankumar* and T.J. Hefley

¹Department of Statistics, Kansas State University, Manhattan, KS 66506, USA.
**meenu@ksu.edu*

In ecology, occupancy data are a contaminated binary response that is used to map the presence or absence of a species. Models for occupancy data are used to estimate the occurrence of a species, where the true presence of a species is a function of a spatially varying process. In the standard spatial occupancy model, most researchers assume that the spatial component is a Gaussian process. This assumption leads to an inability to identify non-traditional spatial dependence, such as discontinuities and abrupt transitions, which are common in ecological data. Machine learning techniques have the potential to identify the non-traditional spatial structure, but these technologies do not account for contamination in the binary response. Machine learning methods were embedded into the Bayesian hierarchical occupancy model to account for non-traditional spatial dependence and contamination in the binary response. A simulation experiment was conducted by selecting a few commonly encountered cases of traditional and non-traditional spatial dependencies in ecology. In the Gibbs sampler, three different machine learning techniques, namely, generalized additive model (GAM), regression trees, and support vector regression (SVR) were considered to estimate the spatial dependence. The results show that embedding the GAM into the Gibbs sampler can capture the smooth transitions of the spatial structure successfully, and embedding the regression tree can identify abrupt transitions and discontinuities in the spatial structure. In all scenarios in the simulation experiment, this approach yields reliable estimates of the model evidence. Moreover, an application of the proposed method was conducted using data on Thomson's gazelle in Tanzania.

Financial assistance from National Science Foundation, USA (DEB-1754491) is acknowledged.

Keywords: Ecological statistics, Bayesian hierarchical models, Occupancy models, Spatial statistics

MEDICAL STATISTICS

Abstract No: 10

TRENDS OF SEXUAL ABUSE IN SRI LANKA

**R.A.S.U. Ranasinghe^{1,*}, S.D. Dharmaratne¹, L.I. Ranasinghe², I.N. Thotawatte¹,
P.G.K. De Silva¹ and W.D.T.S. Welwita¹**

¹*Department of Community Medicine, Faculty of Medicine, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

²*Family Health Bureau, 231 De Saram Place, Colombo 01, Sri Lanka.*

^{*}rasu.ranasinghe@gmail.com

Rate of rape and sexual abuse are on the verge of rising with the urbanization, technological advancements and loss of cultural identities in South East Asia. The data analysis was conducted using the Grave Crime Abstracts of the Department of Police, Sri Lanka from the year 2005. The data of victims less than 16 years were available only from the year 2015. The incidence of rape in the year 2005 was 7.82 per 100,000. It has increased up to the incidence of 8.08 per 100,000 in 2018. The incidence of grave sexual abuse (sexual abuse beside rape) was 2.18 per 100,000 in the year 2005. In 2018, it was 3.47 per 100,000, showing a rise. There were no significant trends according to time in incidences of rape or incidences of grave sexual abuse. Proportions of statutory rape incidence (incidences among women less than 16 years of age) have remained constant from the year 2015 (0.81) to the year 2018 (0.827). Out of them, the proportion of victims who have given their consent for sexual acts have risen from the year 2015 to 2017. But by the year 2018, it has drastically dropped. The perpetrator was known to the victim in more than 95% of both rape and grave sexual abuse cases from the year 2005 to 2018. There is a noteworthy rise in the incidence of rape and sexual offense in Sri Lanka. Underage girls are more willingly becoming victims of rape. In the majority of cases, perpetrators are known by the victims. More research illustrating the reasons and methods of prevention of sexual exploitation is a timely need in Sri Lanka.

Keywords: Grave Sexual Abuse, Rape, Sexual Abuse, Sexual Offense, Statutory Rape.

Abstract No: 11

**MODELING THE DENGUE EPIDEMIOLOGY IN SRI LANKA
USING A RANDOM EFFECT COPULA MODEL**

G.H.S. Karunarathna* and M.R. Sooriyarachchi

Department of Statistics, Faculty of Science, University of Colombo, Colombo 07, Sri Lanka.
*hasani@stat.cmb.ac.lk

Dengue is the most common and widespread mosquito-borne tropical viral disease in Sri Lanka in recent years as several deaths have been reported all over the country. This reveals that analyzing the dengue patients and identify factors associated with the increase of dengue cases will provide supportive documentation for society to reduce the dengue cases. Also, the district clustering effect should be taken into account since dengue is highly associated with climate change which varies between districts. This leads to multilevel modeling for analyzing the dengue patients. Hospital length of stay (LOS) or survival time/competing risk of a dengue patient and platelet count are interesting variables for exploring the relationship between platelet count and several destinations (competing event), such as discharged and dead, among dengue patients. Here, the event is of discrete form and platelet count is in a continuous form. Therefore, analyzing the dengue responses has been carried out by formulating the joint association among a competing event and platelet count under the multilevel scenario. In this paper, we developed a novel joint model for discrete competing risk with continuous outcome as platelet count via Gaussian copula for the dengue epidemiology in Sri Lanka for the period 2006 – 2008 within high incidence 10 districts. This model adopts multilevel utility model formulation of discrete competing risk outcome and multilevel normally distributed explanatory variable. The square root of the platelet count was used as the continuous variable while the survival time with several destination which can be considered as the competing risk was used as discrete variable. Due to convergence issues of the estimation procedure, rescaling was done for the square root of platelet. This model captures the within and between study variability through random effects. The fitted model indicated that time indicators (0-4 days: febrile phase, 4-6 days: critical phase, and 6-10 days: recovery phase), WBC (White Blood Cell), Year, Month, Sex and Classification (DF, DHFI DHFII) are associated factors for the event types and WBC and Year are only factors for the platelet count. Moreover, the study concluded that classification is negatively associated for both event of discharge and dead, and that there is a high possibility to a dead when people are having DHFII. The joint model yields more precise results compared to univariate models.

Keywords: Multilevel models, Competing risk, Utility models, Copula.

Abstract No: 15

**ASSESSMENT OF CORRELATION BETWEEN MAMMOGRAPHIC BREAST DENSITY AND FACTORS AFFECTING BREAST DENSITY:
A QUANTITATIVE APPROACH**

**I.L.U. Chandrasiri^{1,*}, P.B. Hewavithana², U.J.M.A.L. Jayasinghe¹,
P.M.P.C. Gunathilake³ and S.P. Abeywarda³**

¹*Department of Radiography/Radiotherapy, Faculty of Allied Health Sciences,
University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Department of Radiology, Faculty of Medicine, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

³*Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

*lakminiuc@yahoo.com

Breast cancer is the first ranked type of cancer in worldwide which is commonly diagnosed to cause death in females. The breast density plays a considerable role in increasing the risk. As the most commonly utilized imaging modality for breast imaging, mammography provides good quality images of the breast tissues for early detection of the breast cancers. American College of Radiology defines mammographic breast density as the percentage of the content of fibroglandular tissues within the breast. It can be quantified in order to assess the risk of breast cancer due to the increased breast density. The basis of this study was to assess the correlations of the mammographic breast density of the participants with the age, body mass index (BMI) and the duration of breast feeding which were identified as a few factors affecting the breast density. Mammographic breast density for each participant was estimated using a semi-automated computerized programme and the estimated value was used in the statistical analysis along with the collected data during an interview with the participants who were the patients underwent bilateral mammographic examinations at Teaching Hospital Peradeniya. Statistical analysis of the data reveals that the mammographic breast density has a significant negative correlation with the age for all four mammographic projections considered in the study which are right cranio-caudal projection (RCC) ($r = -0.421$, $p = 0.002$), left cranio-caudal projection (LCC) ($r = -0.309$, $p = 0.026$), right mediolateral oblique projection (RMLO) ($r = -0.436$, $p=0.001$) and left mediolateral oblique projection (LMLO) ($r = -0.388$, $p = 0.004$). BMI of the participants showed a significant positive correlation with the mammographic breast density for RCC, LCC, RMLO and LMLO where the Pearson correlation coefficients were 0.559 ($p < 0.001$), 0.444 ($p = 0.001$), 0.459 ($p = 0.001$) and 0.407 ($p = 0.003$), respectively. This analysis did not find any significant correlation between the mammographic breast density and the duration of breast feeding for RCC ($r = 0.014$, $p = 0.923$), LCC ($r = 0.155$, $p = 0.273$), RMLO ($r = -0.003$, $p = 0.983$) and LMLO ($r = 0.075$, $p = 0.599$). Multivariate analysis of variance (MANOVA) showed that the mammographic breast density is significantly higher in pre-menopausal participants than the post-menopausal participants of the study (Wilks' statistics = 0.69, $p < 0.001$). Findings of this study will be useful in predicting the breast density without radiation exposure and to establish a frequency for mammographic screening according to the predicted breast density.

Keywords: Breast cancer, Breast feeding, Mammographic Breast density, Radiation.

Abstract No: 17

IDENTIFYING PRIMARY FACTORS THAT AFFECT THE CHRONIC KIDNEY DISEASE WITH A COMPARATIVE ANALYSIS OF HEAVY METALS IN DRINKING WATER

P.M.S.S. Kumari* and P. Wijekoon

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.
*shamalisujeewa07@gmail.com

The Chronic Kidney Disease (CKD) is an endemic disease scattered in the dry zone of Sri Lanka which causes death either due to a single or multiple effects of chemical constituents or other primary factors from the environment. This study was carried out to determine the possible reasons behind the CKD and delineate possible geo-environmental sources. Data set collected by the National Science Foundation (NSF) of Sri Lanka, and World Health Organization between 2010 and 2012 was used in this study. A population prevalence study was conducted by researchers in Anuradhapura, Badulla and Polonnaruwa (endemic area) and Hambanthota (Non-Endemic) districts where the samples were collected from drinking water sources. A random sample of participants was collected from 220 households. After preprocessing data, only 4902 observations were used for the analysis. Binary logistic model and random forest model were fitted by using population prevalence data to identify the primary factors that affect the CKD. The accuracy of the binary logistic model and the random forest model are 92.98% and 99.91%, respectively, while the predictive abilities of these models are 89.78% and 94.15%. Therefore, the random forest model shows high accuracy and predictive ability compared to the binary logistic model. The results indicated that the chronic illness and approximate age are the highly affected factors to present CKD. Further, the duration of living on an endemic area, source of water storage, gender, source of drinking water and drinking paddy field water can also be identified as the major factors that affect CKD in endemic area. When each of the above significant variables was separately analyzed, it was further identified that there is a high tendency to a CKD present case when all the above variables present. Also, it was found that the males have a higher risk of CKD than females. A statistical comparison test was done to compare the concentrations of heavy metals of drinking water among endemic and non-endemic regions. According to the results, it was observed that there was a significant difference in concentrations of cadmium (Cd) and arsenic (As) in drinking water among the two regions. Although Cd and As concentrations are high in endemic areas compared to the non-endemic areas, there is no significant evidence (p -value > 0.05) to conclude that the concentration of Cd and As exceeds WHO maximum permissible levels.

Keywords: Binary Logistic Model, Chronic Kidney Disease, Random Forest Model, Water Metal Analysis.

Abstract No: 20

ASSESSMENT OF OSTEOPOROSIS RISK USING PREDICTIVE MODELS

B.G.G.K. Somathilake^{1,2,*}, S.D. Viswakula¹ and J.M.K.C. Jeewandara³

¹*Department of Statistics, Faculty of Science, University of Colombo, Colombo 07, Sri Lanka.*

²*National Centre for Primary Care and Allergy Research, University of Sri Jayewardenepura,
Nugegoda 10250, Sri Lanka.*

^{*}gayashaks@gmail.com

Osteoporosis is a silent, progressive bone disorder which deteriorates bone tissues leading to fragility fractures. The quality of life of affected patients could be severely impaired as a result of this chronic disease, which thus intensifies the impact of the physical manifestations of osteoporosis. Hence, the diagnosis of osteoporosis and risk assessment is found to be crucial in clinical practice. Directing a patient to DXA (Dual-energy X-ray Absorptiometry) scanning has to be done wisely since it will lead to the exposure of unnecessary radiation and costs. The paucity of osteoporosis related epidemiology in urban Sri Lanka has been a huge motivation to conduct this particular study among 996 patients who had undergone scanning at a Family Practice Centre in the Western Province. The main objective of the study was to develop two predictive models which could independently assess the osteoporosis risk in male and female patients based on their hip and lumbar spine T-scores. A number of statistical and machine learning techniques such as LASSO Ordinal Logistic Regression, Bagging, Random Forest, Boosting, Boosted Random Forest, Ordinal Forest and Support Vector Machines were utilized in order to achieve these optimal models by comparing the measures of prediction accuracy under the concept of parsimony. Boosted random forest algorithm was resulted as optimal with 56.10% accuracy for the postmenopausal and perimenopausal women while for men, it was the boosting algorithm which produced 61.54% accuracy. Among the clinical risk factors identified through the study, the advancing age and lower body mass index of the patients were observed as the highly influential factors of osteoporosis risk regardless of their gender. Moreover, the menopausal status of women and waist circumference of men were also resulted as significant from the developed models. Cluster Analysis was applied in order to identify if there exists a different grouping scheme based on the T-scores with regard to the state of osteoporosis instead of the usual categories normal, osteopenia and osteoporosis using the T-scores. The results indicated the clear existence of distinct clusters in both male and female groups. This suggests the possibility for the existence of a new classification criterion which could be more appropriate for the Sri Lankan community in diagnosing patients. The study concludes that it is possible to propose a strategy based on the models which would prompt initiation of treatment in patients with very high risk while limiting scanning in those with low risk in order to resolve the problem of expensive scanning, which then will not incur any unnecessary risk of radiation or costs. Hence, this would clearly do justice in the utilization of resources in diagnosing and assessing the risk of osteoporosis in Sri Lanka.

Keywords: Machine Learning, Osteoporosis, Predictive Modeling, Risk Assessment, Risk Factors.

Abstract No: 22

**MODEL-BASED APPROACH TO OPTIMIZE CLINICAL OUTCOMES IN
NEONATAL ABSTINENCE SYNDROME USING REAL WORLD DATA**

N. Wijekoon^{1,*}, O. Aduroja², J. M. Biggs³, D. El-Metwally² and M. Gopalakrishnan⁴

¹*Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA*

²*Department of Pediatrics, School of Medicine, University of Maryland, Baltimore, MD 21250, USA*

³*University of Maryland Medical Center, University of Maryland, Baltimore, MD 21250, USA.*

⁴*Center for Translational Medicine, School of Pharmacy, University of Maryland, Baltimore, MD 21250, USA.*

**nadeesri.wijekoon@umbc.edu*

Neonatal abstinence syndrome (NAS) is a drug withdrawal syndrome that neonates exposed to opioids *in-utero* may experience after birth. At least 75% of neonates with NAS require treatment, with morphine being most commonly used drug. Currently, the morphine dosing adjustments are often empiric or stepwise in nature; associated with longer hospital stays. The aim of the study is to use a real world data-driven approach to optimize morphine dosing in neonates with NAS to improve clinical outcomes. Longitudinal morphine dose and clinical response [Modified Finnegan Score (MFS)] data along with maternal and infant baseline factors were collected using a retrospective cohort design from the electronic medical records of infants with NAS ($N = 189$) admitted to the Neonatal Intensive Care Unit at the University of Maryland Medical Center (2013-2017). A dynamic linear mixed effects (DLME) model with independent covariance structure was used to develop the relationship between MFS and morphine dose adjusting for baseline risk factors. Simulation based model evaluation was performed by comparing the observed and predicted using clinically meaningful metrics such as time on treatment. Maternal methadone use, poly-substance drugs, race, Benzodiazepine exposure and previous morphine dose were significant predictors of MFS. Significant positive autocorrelations of previous two MFS with current MFS was observed ($\rho_1 = 0.56$ and $\rho_2 = 0.13$ respectively). On an average, for a 100 mcg increase in morphine dose, the MFS decreased by 0.5 units. The model evaluation showed that observed and predicted time on treatment (median: 11.0 vs. 9.8 days) was not significantly different (p -value = 0.28). A model-based framework was developed to describe the MFS–morphine dose relationship using real world data. Further improvements to the DLME model are underway to develop an adaptive, precision dosing strategy for infants with NAS that could lead to better clinical outcomes.

Keywords: Dynamic Linear Mixed Model, Modified Finnegan Score, Neonatal Abstinence syndrome, NAS.

Abstract No: 26

PREDICTION OF AGE FROM ERUPTION STATUS OF PERMANENT TEETH IN SRI LANKAN CHILDREN

B.D.H.H. De Silva¹, L.S. Nawarathna^{1,*} and V.S.N. Vithanaarachchi²

¹*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

²*Department of Community Dental Health, Faculty of Dental Sciences, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

^{*}lakshikas@pdn.ac.lk

Prediction of age is a predominant facet in forensic and clinical fields. Forensic odontology is used to predict the age, by using permanent teeth which are resistant to high temperature and any mass disaster than other parts of the body. Although many studies have been carried out in other countries, this is the first study of age prediction using eruption status of permanent teeth for Sri Lankans. Age-related memory loss, memory loss associated with dementia and absence of any official documents to verify their age are some of the main reasons why people have no knowledge about their age. Therefore, age prediction finds its use in various situations, such as identification, admission purposes, employment, criminal issues and judicial punishments. The main objective of this study was to predict the age of a child using the eruption status of permanent teeth. Eruption time difference between both genders (females and males) and both sides (left and right) were also analyzed teeth wise. This cross-sectional study was conducted on 3321 individuals (1681 males and 1640 females) and eruption status of permanent teeth was noted in the age group from 4 to 21 years. Multiple linear regression (MLR), classification and regression trees (CART) algorithm, gradient boosting (GB) classifier and extreme gradient boost (XGBoost) classifier were used to predict the age. Results were validated using 10-fold cross-validation technique. Root mean squared error (RMSE), mean absolute error (MAE) and R-squared values were used for selecting the best model. A statistically significant difference in eruption time was found between males and females of all permanent teeth, except for central incisor and second premolar. No statistical difference was found between eruption time of the left and right side. Upper left central incisor, lateral incisor, canine, first premolar, second premolar, upper right lateral incisor, canine, first premolar, second premolar, second molar, lower left lateral incisor, second molar and lower right lateral incisor, canine, first premolar, second premolar were the significant variables of the fitted models for predicting the age. The best model for age prediction was the extreme gradient boosting (XGBoost) model which gives the highest accuracy (82%). Therefore, the extreme gradient boosting model can be used to predict the age using eruption status of permanent teeth.

Keywords: Classification, Gradient Boosting Classifier, Permanent Teeth, Regression Trees, Tooth Eruption.

Abstract No: 29

MULTI-CLASS CLASSIFICATION APPROACH TO IDENTIFY THE STAGE OF AUTISM SPECTRUM DISORDER

M. Kayanan^{1,2,*} and P. Wijekoon^{1,3}

¹*Postgraduate Institute of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya 43000, Sri Lanka,*

³*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

*mgayanan@vau.jfn.ac.lk

Autism Spectrum Disorder (ASD) is known as a neurodevelopmental disorder that affects communication, social interaction and behavioural skills. ASD affects children at the age of two years old and continues lifelong. Medicine cannot cure ASD, but early interventions can be helpful to reduce adverse effects of ASD. Since ASD is a spectrum disorder, it can be classified as mild, moderate and severe stages. Based on the ASD stage, the appropriate therapies can be prescribed to an ASD diagnosed child. Nowadays, screening methods such as Screening Tool for Autism in Toddlers and Young Children (STAT), Childhood Autism Rating Scale (CARS-2), and Autism Spectrum Quotient (AQ), have been used to diagnose the ASD. For efficient decision making in ASD diagnoses, the machine learning techniques, namely automated classification methods, have been used by many researchers in recent literature. In this research, multiclass classification techniques, such as Ordinal Logistic regression, Decision tree, Conditional Inference Trees and Random Forest were used to identify the stage of ASD. Further, the prediction accuracy of these methods was evaluated using real ASD dataset. The analysis reveals that Ordinal Logistic regression approach provides the best results with very high accuracy for both training and test data. Due to the rapid increase of ASD, early diagnosis of stage of ASD with the support of classification models will undoubtedly contribute to a greater extent in decision making.

Keywords: Autism Spectrum Disorder, ASD Stage, Machine learning, Multiclass Classification.

Abstract No: 30

EARLY RECOGNITION OF PREMATURE BIRTH: A COMPARATIVE ANALYSIS

T.M.S. Priyankara¹ and S.P. Abeysundara^{2,*}

¹*National Institute of Fundamental Studies, Hantana Road, Hantana, Kandy 20000, Sri Lanka.*

²*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

^{*}*sachitha@pdn.ac.lk*

Premature birth is one of the severe health issues which occurs due to babies born before 37th week of gestation. Specific courses for premature birth are still not have been identified. Early detecting may improve effective clinical approaches to reduce the effect on both infants and maternal health. It has been identified that uterine electromyography (EMG) signals recorded from the abdominal surface have electrical activities corresponds to some physical activities of uterus and have a discriminate ability for preterm records from others. In this study, a set of uterine EMG records were used for investigating effective feature extraction techniques for predicting preterm birth using an open data set which consists of 38 preterm and 262 term EMG uterine signals. Combination of EMG signal feature extraction techniques which have been used in previous studies were considered and used for each uterine EMG signal which were recorded from different filtering parameters and with different electrode positions on abdominal surface. Logistic regression, Least Absolute Shrinkage and Selection Operator (LASSO) classifier, Decision tree classifier and Support vector machine classifiers were used as classifications algorithms with different combination of extracted feature sets with and without Combination of Synthetic Minority Over-sampling Technique (SMOTE) and under sampling technique and dimension reduction techniques. Even though all classifiers gave significant results, support vector machine classifier with SMOTE technique and principal component analysis technique gave the highest classification performance of area under curve (AUC) that is 94.87%.

Keywords: Feature Extraction, Preterm Birth, Gestation, SMOTE, Uterine EMG Signals.

Abstract No: 31

SERUM TGF BETA 1 AS AN EARLY MARKER OF CHRONIC KIDNEY DISEASE OF UNCERTAIN AETIOLOGY

**W.B.N.T. Fernando^{1,*}, K.R.D. Hemage¹, M.Z. Badurdeen¹, T.W. Hettiarachchi¹,
H.T.K. Abeysundara² and N. Nanayakkara³**

¹*Center for Education, Research & Training on Kidney Diseases (CERTKiD), Faculty of Medicine,
University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

³*Transplant and Dialysis Unit, Teaching Hospital, Kandy 20000, Sri Lanka.
buddhifernando08@gmail.com

Chronic Kidney Disease of uncertain aetiology (CKDu) is a recently described tubular interstitial nephropathy in some tropical countries including Sri Lanka. The major limitation of the traditional markers is the inability to detect the disease at early stages. The aim of this study was to identify the importance of the serum Transforming Growth Factor 1 (TGF beta1) as an early marker for CKDu in Sri Lanka. In the case group, 76 of definite, non-dialysis CKDu patients were recruited. For control groups, endemic control (EC 79), endemic CKD (ECKD 79), non-endemic control (NEC 85) and non-endemic CKD (NECKD 85) patients were recruited. Serum TGF beta 1 was measured in a Luminex MAGPIX platform. The higher mean value of TGF beta1 (69587 pg/mL) was obtained from the CKDu group compared to other control groups. There was a significant difference in mean TGF beta1 of CKDu with the other four control groups ($p < 0.001$). A significant difference was seen in CKDu group between stages 2 and 4, 3a and 4 and 3b and 4. In CKDu and NECKD groups, there was a negative correlation with serum creatinine ($r = -0.293$ and -0.247) and positive correlation ($r = 0.332$ and 0.242) with the eGFR which were significant ($p < 0.01$). According to the ROC analysis, to differentiate CKDu from the other endemic and non-endemic CKD forms, TGF beta1 showed a good area under the curve AUC than the serum creatinine (0.955 and 0.987 respectively). For healthy control groups, TGF beta1 showed a slightly decreased value of AUC than the serum creatinine (0.862 and 0.902). The negative correlation of TGF beta1 with the serum creatinine also supports the feature of early detection of the CKDu. The results suggest that serum TGF beta1 as a marker of early detection of CKDu and useful to differentiate CKDu from other CKD types.

Financial assistance from National Research Council of Sri Lanka (Grant No: TO 14-05) is acknowledged.

Keywords: Chronic Kidney Disease of uncertain aetiology, Early marker, Transforming Growth Factor beta 1, Tubular Interstitial Nephropathy.

Abstract No: 35

SEX DETERMINATION USING THE HUMERUS OF SKELETONS OF CONTEMPORARY SRI LANKAN POPULATION

T. Selvamalai¹, L.S. Nawarathna^{2,*} and D. Nanayakkara³

¹*Department of Economics and Management, Vavuniya Campus of the University of Jaffna, Vavuniya 43000, Sri Lanka.*

²*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

³*Division of Anatomy, Department of Basic Sciences, Faculty of Dental Sciences, University of Peradeniya, Peradeniya 20400, Sri Lanka.
lakshikas@pdn.ac.lk

Unidentified human remains create various problems at both the legal and emotional levels of people. Sex determination is an essential step for the identification when a decomposed body is recovered. It is essential to determine the sex of unknown remains of human skeletons to speed up the investigations of the forensic and archeological field as well. The skull and some postcranial elements, such as the humerus, femur and tibia, have been used in sex determination. Hence, the study decided to determine the sex using humerus which is a long bone of the upper extremity. Besides, the study also analysed the significant differences between male and female measurements. For the study, the dimensions of maximum length, intercondylar breadth, the vertical and transverse diameter of left side humeri were measured in millimeters (mm) of 33 skeletons of adult Sri Lankans (23 male and 10 female) acquired from the Department of Basic Sciences, Faculty of Dental Sciences, University of Peradeniya, Sri Lanka. The study selected univariate and multivariate quadratic discriminant function due to the presence of heterogeneity in variances along with pre-conditions of both normality and unequal means between males and females. Shapiro-Wilk and Mardia's tests were used to test the univariate and multivariate normality respectively. The unequal mean and equal variance between males and females were tested using Welch two-sample t-test and the Box-M test, respectively. The sizes of maximum length, intercondylar breadth, the vertical and transverse diameter of the head of the humerus were significantly larger in men than in females. The univariate quadratic discriminant function with the transverse diameter of the head of the humerus is adequate to determine the sex with 100% accuracy in contrast to the multivariate model with maximum length, vertical and transverse diameter of the head of humerus which provided slightly less accuracy (97%) when both original and cross-validation data were used to test the accuracy. Further, in this study, the maximum length (78%) and intercondylar breadth (85%) are the dimensions showing lowest sexual dimorphism compared to the transverse and vertical diameter of the head of the humerus. Finally, the study concludes that the transverse diameter of the head of the humerus is the best among the humeral measurements to predict the sex correctly.

Keywords: Humerus Dimensions, Quadratic Discriminant Function, Sex Determination.

Abstract No: 37

**DETERMINATION OF SEX USING FOOT, HAND AND HAND PRINT
MEASUREMENTS IN A SRI LANKAN POPULATION**

A.B.S.N. Dayarathne¹, L. S. Nawarathna^{1,*} and D. Nanayakkara²

¹*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

²*Division of Anatomy, Department of Basic Sciences, Faculty of Dental Sciences,
University of Peradeniya, Peradeniya 20400, Sri Lanka
lakshikas@pdn.ac.lk

Foot, footprint, hand and handprint measurements provide excellent background in the process of individual identification in the fields of forensic science, archaeology and medical studies. Also, these measurements are widely used in forensic science to resolve identification difficulties of mutilated and dismembered bodies. Since sex determination covers one of the prior identification requirements of a victim, many studies have been undertaken in this aspect in various populations and have shown that reliable results can be obtained using arms, legs and fingers. This study is focused on determining the sex using foot, footprint, hand and handprint data of the Sinhalese population in Sri Lanka. The objective of this study was to identify the differences of those measurements in males and females to build models to predict the sex. The dataset consists of 117 individuals (51 males and 66 females) in age range of 20 to 28 years. Using standard techniques and instruments ten measurements of foot and footprint (foot length, foot breadth, foot heel breadth, footprint breadth, footprint heel breadth and five-toe lengths) and eight measurements of hand and handprint (hand length, handbreadth, palm length and five finger lengths) were recorded. Students t-test was applied for identifying differences between males and females and found that means of all measurements were significantly larger in males (p-value < 0.05). Paired t-test was applied to check for bilateral asymmetries and it was found that foot length, thumb and smallfinger lengths were significantly different in males and females. The principal component analysis was applied to avoid multicollinearity and found that eleven variables of the foot were loaded into three principal components and for each hand and handprint, all eight variables were loaded into two principal components. Support Vector Machine (SVM) was applied for classification of sex and was found that the highest accuracy is given by foot measurements as 93.75% with all predictor variables while handbreadth estimates sex with 93.75% of accuracy. The results were validated using cross-validation techniques. The results of this study showed that foot, footprint, hand and handprint measurements can be used to predict the sex for the identification in forensic science and also, the outcomes of this study will be useful in regenerating crime scenes or robberies and also to find missing persons.

Keywords: Bilateral Asymmetry, Classification, Forensic Science, Principal Component Analysis, Sex determination , Support Vector Machine.

Abstract No: 42

**PREVALENCE OF HYPERURICEMIA AND ANAEMIA AMONG PATIENTS
WITH CHRONIC KIDNEY DISEASE OF UNKNOWN AETIOLOGY IN
WILGAMUWA, SRI LANKA**

**H.P.M. Hewavitharane¹, N. Nanayakkara^{2,*}, S. Anand³, R. Chandrajith⁴, P. Vlahos⁵,
S. Schensul⁶, J. Schensul⁶, L. Haider⁶ and K.T. Silva⁷**

¹*Center for Education Research and Training on Kidney Diseases, Faculty of Medicine,
University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Nephrology and Kidney Transplant Unit, National Hospital, Kandy 20000, Sri Lanka.*

³*Medical Centre, Stanford University, CA 94305, USA.*

⁴*Department of Geology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.*

⁵*Department of Marine Sciences, University of Connecticut, Storrs, CT 06269, USA.*

⁶*Health Centre, University of Connecticut, Storrs, CT 06269, USA.*

⁷*Department of Sociology, Faculty of Arts, University of Peradeniya, Peradeniya 20400, Sri Lanka.*

^{*}*nishantha4313@gmail.com*

Chronic Kidney Disease of unknown aetiology (CKDu) is a progressive disease that causes irreversible renal failure and eventually the death. The causative factors for CKDu are yet to be conclusively proven. Sri Lanka is one of the countries affected by CKDu and it has become a severe health issue among the farmers in the dry zone. CKD/CKDu progresses via five stages and symptoms become more visible at the advanced stages (i.e. stage 4 and 5). The progression of the disease can be monitored by periodic analysis of biochemical parameters such as serum creatinine, haemoglobin, uric acid and electrolytes. Hyperuricemia (uric acid > 7 mg/dl) and anaemia (haemoglobin < 12 mg/dl) are characteristic clinical features in CKD patients. The main objective of this study was to assess the prevalence of hyperuricemia and anaemia among the stage 3 CKDu patients in Wilgamuwa, a recently identified CKDu hotspot in Mathale district, Central Province, Sri Lanka. A descriptive cross sectional study was carried out to identify the socio demographic features and to assess the prevalence of hyperuricemia and anaemia in selected CKDu population. The selected CKDu population was comprised of 302 CKDu stage 3 (i.e. eGFR 30-60 ml/min/1.73 m²) patients registered at the renal clinic in Wilgamuwa regional hospital. An interviewer administered questionnaire was used to evaluate the demographic features. Biochemical parameters were obtained from the clinic records. Majority (76%) of patients were males indicating the usual CKDu distribution of male: female 3:1. Mean age was 53 years. Majority of patient's body mass index (BMI) was within the normal range (18.5-24.5 kg/m²). Moreover, 76% of females and 68% of males were anaemic and 62% and 64% of male and female patients respectively had elevated uric acid levels (uric acid > 7 mg/dl). Farming behaviour, alcohol and tobacco usage and water drinking behaviour were also assessed. The findings of this study indicate a strong relationship between the CKDu occurrence and hyperuricemia and anaemia among the patients in Wilgamuwa region. Association between Hyperuricemia and anaemia to the progression of CKDu to be assessed.

Financial assistance from Fogarty International Center, USA. (Grant No R21TW010425) is acknowledged.

Keywords: Anaemia, CKDu, Hyperuricemia, Progression.

Abstract No: 43

**TRACE ELEMENTS LEVELS IN RICE CONSUMED BY PATIENTS
WITH TUBULITIS**

**O.S.S. Preamrathne^{1,*}, R. Chandrajith², N. Nanayakkara³, H.P.M. Hewavitharane¹,
S. Wijethunge⁴, R.D. Yapa⁵, H.T.K. Abeysundara⁵, N. Elladeniya³,
D.P.G.S. Kumudumali¹, K.P.S. Madushan³ and K.O.C.U. Samarasiri³**

¹*Centre for Education Research and Training on Kidney Diseases, Faculty of Medicine,
University of Peradeniya, Peradeniya 20400, Sri Lanka.*

²*Department of Geology, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

³*Nephrology and Kidney Transplant unit, National Hospital, Kandy 20000, Sri Lanka*

⁴*Department of Pathology, Faculty of Medicine, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

⁵*Department of Statistics and Computer Sciences, Faculty of Science, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

*premarathneshaki@gmail.com

The accumulation of toxic heavy metals in the human body has been considered as a cause for renal dysfunction. Symptomatic Chronic Kidney Disease of uncertain aetiology (CKDu) with tubulitis is a recently described clinical entity in Sri Lanka. The patients presented with acute or subacute syndrome of fever, myalgia, back pain, dysuria with elevated serum creatinine. This acute illness is hypothesized due to recent exposure to risk behaviours, nephrotoxin or both. Contaminated food is a primary pathway for human exposure to environmental toxins. Rice is the second most prevalent cereal crop in the world and is the staple food of most Asian countries; Hence trace elemental analysis is important to determine the intake of trace elements. The objective of this study was to determine the exposure to heavy metals via food chain. Thirty two rice samples were collected from 30 biopsy proven patients with tubulitis, selected from Girandurukotte, Wilgamuwa and Kandy renal clinics. Rice samples were digested and analysed for 45 trace elements. Mean value of Pb was 0.0012 mg/kg, ranging from (0.0050 – 0.0043) mg/kg. The mean value of As was detected < 0.05 mg/kg, with zero ranging values. The mean Cd content was 0.0060 mg/kg. Mean values of Fe, Ni Cu, Zn were 0.2926 mg/kg, 0.0127 mg/kg, 0.0097 mg/kg and 0.0760 mg/kg, respectively while the other toxic element Hg was always < 0.0001 mg/kg. All these levels were less than the CODEX recommended levels. Findings do not support the possibility of consumption of heavy metal contaminated rice as the cause of CKDu with tubulitis.

Financial assistance from National Science Foundation. (Grant No RPHS/2016/CKDu/06) is acknowledged.

Keywords: CKDu, Trace Elements, Tubulitis.

Abstract No: 44

**PARTICIPATION OF SPORT IMPROVING ACADEMIC ACHIEVEMENTS,
HEALTHY LIVING AND WELL-BEING**

L. Liyanage¹ and R.P. Premaraja^{2,*}

¹*School of Computing, Engineering and Mathematics, Western Sydney University,
Locked Bag 1797, Penrith NSW 2751, Australia.*

²*The Open University of Sri Lanka, Nawala, Nugegoda 10250. Sri Lanka.
fr.ruwan@yahoo.com

This study investigates the distinct influence that sports activities contribute towards the development of the cognitive nature of the individuals and the association of sports activity and physical development resulting in an improvement of cognitive ability. Term marks for all subjects are collected using a questionnaire from a cohort of seventy-eight students across grades 6 to 11. Findings reveal that the students who engage in sports practices one day per week have always performed better than students who did not participate in sport. Students who participate in two or three days per week also performed better on many occasions. However, the test of significance reveals that there is a significant improvement in learning languages, Sinhala and English, where cognitive skills are needed when students participate in sports. When learning Mathematics and Science where critical thinking is needed a significant improvement was not supported. Finally, a case study is carried out to design a qualitative methodological approach to be used for future research in this topic to explore associations of sports activities and academic achievements. In this study, data is gathered from semi-structured interviews from a small cohort of students who excel in sports having achieved a win/victory in provincial level championship games and at the national level. Focus group discussions of the class teachers and parents of these students are then carried out to investigate those students' perceptions of academic achievements and their study commitments at home. These students have followed five common as well as some inherent study techniques and strategies on their learning to foster academic success. Most significant being enthusiasm in classroom activities and studying according to a time table respectively. A more extensive study of this nature will be significant to students, teachers, curriculum planners, parents and educators, school administrators as this will provide empirical evidence of the association of sports and academic achievements and wellbeing for making informed judgments and decisions.

Keywords: Academic Achievements, Experimental Design, Health and Wellbeing, School Curriculum, Sport.

Abstract No: 45

**LATENT RELATIONSHIPS AMONG DEMOGRAPHIC, SOCIO-ECONOMIC,
ANTHROPOMETRIC AND SERUM BIOCHEMICAL CHARACTERISTICS IN A
POPULATION ON DIETARY ADVICE: MULTIPLE CORRESPONDENCE
ANALYSIS APPROACH**

G.G.S. Ranasinghe^{1,*} and S. Samita²

¹*Postgraduate Institute of Agriculture, No 55, Old Galaha Road, Peradeniya 20400, Sri Lanka.*

²*Department of Crop Science, Faculty of Agriculture, University of Peradeniya,
Peradeniya 20400, Sri Lanka.*

^{*}*sachini.ranasighe91@gmail.com*

There is a large number of multi-category variables which are risk factors for Non-Communicable Diseases (NCD). It is a challenge to detect relationships among the levels of variables and the risk of diseases. Moreover, detecting risk factors of NCDs including heart diseases are crucial, so that screening should be done precisely. The aim of this study was to identify unobserved relationships among demographic, socio-economic, anthropometric and serum biochemical characteristics in a population on dietary advice. The sample consisted of 368 women who attended a dietitian's clinic at National Hospital of Sri Lanka and the data were collected on 130 variables. Since latent relationships between levels of the variables were the main focus, standard multivariate techniques based on multivariate normal distribution cannot be used for the analysis. Thus, Multiple Correspondence Analysis (MCA) was used to detect latent relationships. Out of 130 variables, 22 variables were chosen for initial screening then 5 variables were identified for MCA, namely, Low-Density Lipoprotein (LDL) level, level of education, family history of hypertension, moderate-intensity physical activity level and ethnicity. Using the inertia values from the analysis of the indicator matrix, 5 components were identified, and they explained 71.12% of the total observed variability. Each component was contributed by a combination of different levels of variables. Burt table values and column contributions table were used to identify latent relationships. Levels of the variables, LDL level, ethnicity, and level of education had a significant contribution for relationships identified. Individuals with secondary education, borderline LDL level, family history of hypertension and adequate physical activity was one of the relationships identified. Two associations between ethnicity and LDL level were also identified. The first was high LDL level with Tamil ethnicity and the second was normal LDL level with Sinhalese. Moreover, the level of education was negatively associated with the risk of hypertension. The relationships identified among levels of different variables could effectively be used, especially in developing awareness programs on NCDs. When compared with other multivariate techniques, MCA is less biased and it is minimally affected by the noise in a dataset. In addition, MCA was found to be useful in detecting relationships among the levels of variables, hence it can be used to identify latent relationships between levels of potential risk factors and diseases.

Keywords: Hypertension, Low-Density Lipoprotein, Multiple correspondence analysis, Multivariate methods, Non-communicable diseases.

ORGANIZING COMMITTEE

LOCAL ADVISORY COMMITTEE

- Dean, Faculty of Science, University of Peradeniya
- Director, Postgraduate Institute of Science, University of Peradeniya
- Heads of Departments, Faculty of Science, University of Peradeniya
- Chairman, Faculty Research Committee, Faculty of Science, University of Peradeniya
- Director, International Affairs Office, University of Peradeniya

INTERNATIONAL ADVISORY COMMITTEE

- Prof. Bimal Sinha, Co-Chairperson, Department of Mathematics & Statistics, University of Maryland Baltimore County
- Prof. Thomas Mathew, Department of Mathematics & Statistics, University of Maryland Baltimore County
- Prof. Nagaraj Neerchal, Department of Mathematics & Statistics, University of Maryland Baltimore County

MAIN ORGANIZING COMMITTEE

- Prof. P. Wijekoon, Conference Co-chairperson
- Dr. Lakshika S. Nawarathna, Conference Secretary
- Prof. K.W.G. Rekha Nianthi, Department of Geography, Faculty of Arts, University of Peradeniya
- Dr. U.J.M.A.L. Jayasinghe, Department of Radiography/Radiotherapy, Faculty of Allied Health Sciences, University of Peradeniya
- Dr. Niluka Gunawardana, Department of Basic Sciences, Faculty of Dental Sciences, University of Peradeniya
- Prof. P.V.R. Kumarasiri, Department of Community Medicine, Faculty of Medicine, University of Peradeniya
- Conveners of Sub Committees

EDITORIAL AND TECHNICAL PROGRAMME SUB-COMMITTEE

- Dr. Sachith Abeysundara (Convener)
- Prof. Namal Priyantha
- Prof. W.A.I.P. Karunaratna
- Prof. P. Wijekoon
- Dr. Lakshika S. Nawarathna

CEREMONIAL SUB-COMMITTEE

- Dr. Hemalika Abeysundara (Convener)
- Dr. N.C. Bandara
- Dr. Samodha Subhashini Gunthilake
- Dr. B.M.K. Premasiri
- Dr. Nadeesha Koralagedara

FOOD SUB-COMMITTEE

- Dr. Ruwanthini Siyambalapitiya (Convener)
- Dr. Oshani Erunika Dayaratna (Convener)
- Dr. Lahiru Narangammamana
- Ms. Buddhi Nanayakkara
- Dr. Piyumali K. Perera
- Mr. Saumya Bandara

FUND RAISING SUB-COMMITTEE

- Prof. Dharshana Yapa (Convener)
- Dr. Amalka Pinidhaarachchi
- Director, Postgraduate Institute of Science, University of Peradeniya

LOGISTICS SUB-COMMITTEE

- Dr. Hakim Usoof (Convener)
- Dr. Priyanga Wijesinghe
- Dr. Suranjith Gama-Arachchige
- Dr. R.J.K.U. Ranatunga
- Dr. Chandima Ariyaratne

MEDIA-PROGRAMMING-PUBLICITY SUB-COMMITTEE

- Mr. Prabath Gunathilake (Convener)
- Dr. Prassanna Liyanage
- Dr. Ashwini Amarasinghe

REGISTRATION SUB-COMMITTEE

- Dr. Mahasen Dehideniya (Convener)
- Dr. W.A.P.P. de Silva
- Dr. Rekha de Silva
- Dr. Dimanthi Uduwela
- Dr. Anuradhi Jayaweera

SPONSORS

Gold Sponsors:

- People's Bank, Peradeniya, Sri Lanka
- International Affairs Office, University of Peradeniya, Sri Lanka

Silver Sponsor:

- Sri Lanka-India Foundation

Bronze Sponsor:

- Wisewel LLC, Cleveland, Ohio, United States

Other Sponsors:

- Bank of Ceylon, Peradeniya, Sri Lanka
- English Tea Shop

AUTHOR INDEX

A

- Abeysondara, H.T.K. 50, 54
 Abeysondara, S.P. 43, 49
 Adikari, N. 13
 Aduroja, O.. 46
 Anand, S. 53
 Arachchi, M.A.O.D.M. 28
 Arambawatta, A.K.S. 17

B

- Badurdeen, M.Z. 50
 Balasubramanya, S. 22, 23
 Biggs, J.M.. 46
 Biswas, S. 11
 Budisic, M. 10

C

- Chandana, M.K.L. 34
 Chandrajith, R. 53, 54
 Chandrasiri, I.L.U. 43
 Chathuranga, H.N. 34
 Choudhary, P.K. 7

D

- Dayarathne, A.B.S.N. 52
 Devaisy, S. 28, 31
 Dharmarathna, G.D.P. 36
 Dharmaratne, S.D. 41
 Dissanayake, D. 16
 Dissanayake, D.P.V.R. 25

E

- El-Metwally, D. 46
 Elladeniya, N. 54

F

- Fernando, W.B.N.T. 50

G

- Gopalakrishnan, M.. 46
 Gunathilake, P.M.P.C. 43

H

- Haider, L. 53
 Hapugoda, J.C. 14
 Hefley, T.J. 37
 Hemage, K.R.D. 50
 Hettiarachchi, T.W. 50
 Hewavithana, P.B. 43
 Hewavitharane, H.P.M. 53, 54
 Horbulyk, T. 22, 23

J

- Jayasinghe, U.J.M.A.L. 43
 Jayasundara, U.K. 35
 Jayaweera, J.A.A. 17
 Jeewandara, J.M.K.C. 45
 Jeganathan, P. 9

K

- Kafle, K. 22, 23
 Karunarathna, G.H.S. 15, 42
 Karunasinghe, D.S.K. 32
 Kayanan, M. 48
 Kumari, H.M.N.S. 26
 Kumari, P.M.S.S. 44
 Kumudumali, D.P.G.S. 54

L

- Liyanage, L. 8, 30, 55
 Lufkin, L. 10

M

- Madushan, K.P.S. 54
 Manthrithilake, H. 22, 23
 Mathew, T. 2
 Mohankumar, N.M. 37
 Mondal, S. 10

N

- Nanayakkara, C.D. 17
 Nanayakkara, D. 51, 52
 Nanayakkara, N. 50, 53, 54
 Nandy, K. 12
 Nandy, R.R. 5

Navaratna, W.C.W.	36
Nawarathna, L.S.	47, 51, 52
Nayak, T.	3
Neerchal, N.K.	4
Nianthi, K.W.G.R.	16, 33
Nisansala, S.M.J.H.	29
Nussbaum, B.D.	1
 P	
Pathiraja, B.K.	35
Peiris, H.R.D.	17
Peiris, T.S.G.	21, 27
Preamrathne, O.S.S.	54
Premaraja, R.P.	55
Priyankara, T.M.S.	49
 R	
Rajapaksha, R.P.G.	33
Ranasinghe, G.G.S.	56
Ranasinghe, L.I.	41
Ranasinghe, R.A.S.U.	41
Rankothgedara, L.N.K.	7
 S	
Sakunthala, K.A.U.	21
Samarasiri, K.O.C.U.	54
Samita, S.	56
Schensul, J.	53
Schensul, S.	53
Selvamalai, T.	51
Selvaratnam, A.	24
Senanayake, F.R.	34
Silva, B.D.H.H.D.	47
Silva, H.P.T.N.	27
Silva, K.T.	53
Silva, P.G.K.D.	41
Somarathne, E.D.T.	32
Somathilake, B.G.G.K.	45
Sooriyarachchi, M.R.	13, 14, 15, 42
Stifel, D.	22, 23
Subasinghe, S.I.S.	33
Sur, S.	6, 10
 T	
Thotawatte, I.N.	41
Tillekeratne, L.K.C.	36
 U	
Uduwarage, U.D.T.M.	31
 V	
Vishwavidyapeeth, C.	4
Viswakula, S.D.	45
Vithanaarachchi, V.S.N.	47
Vlahos, P.	53
 W	
Weliwita, W.D.T.S.	41
Wijayakulasooriya, J.V.	32
Wijekoon, N.	46
Wijekoon, P.	24, 25, 26, 29, 44, 48
Wijesekara, W.M.L.K.N.	30
Wijethunge, S.	54
 Y	
Yapa, R.D.	54
Yuan, X.	11
 Z	
Zoysa, D.D.	34