

Special
Collection

Biocompatible Multifunctional Theranostic Nanoprobe for Imaging and Chemotherapy in Solid-Tumor-Bearing Mice

Chandrasekharan Rajalekshmi Dhanya^{+, [a]}, Jaishree Jeyaraman^{+, [b]}, Sherin Sainulabdeen,^[a] Mohanannair S. Soumya,^[a] Annie Abraham,^{*[a]} and Sri Sivakumar^{*[b]}

Theranostic carriers are widely studied as they can be exploited for both imaging and drug delivery. However, although there are reports suggesting their *in vitro* level imaging and drug release, there is necessity to display the application (drug delivery/bioimaging/toxicity) at an *in vivo* level, which is a prerequisite for clinical use. Herein, we demonstrate targeted *in vivo* imaging and biodistribution in Swiss Albino mice solid tumor model using PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles (concentration of nanoparticles ~3.4 μM LaVO₄/kg body weight). The capsules were further loaded with doxorubicin for drug delivery which shows tumor regression at different time intervals in tumor mice model. We

have further investigated the tumor marker enzymes including β-glucuronidase, myeloperoxidase, lactate dehydrogenase, and alkaline phosphatase which clearly suggest the reversion to near normal levels after treating with doxorubicin-loaded polymer capsules for 30 days. comet assay shows DNA damage in tumor cells induced by doxorubicin. Histology studies performed in tumor tissue and liver show obliteration of tumor cells after treating with doxorubicin-loaded PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles. It has also been observed that the weight of the spleen which was enlarged in solid-tumor-bearing mice is significantly lower in animals treated with drug-loaded capsules.

1. Introduction

Development of theranostic vehicles is one of the propelling areas in nanobiotechnology because of their multifunctionality (diagnosis and therapy). Theranostic vehicle may possess one or multiple diagnosis probe (e.g. fluorescence, magnetic, X-ray etc.) along with therapeutic molecule (e.g. drugs, proteins, genes etc.). Many nanomaterials-based theranostic agents such as metallic^[1,2] (e.g. gold, silver),^[2] inorganic (e.g. iron oxide, lanthanide ion doped nanoparticles),^[3–11] polymeric^[12] (e.g. polylactone,^[13] carboxylated poly [styrene-co-chloromethyl styrene]-graft-poly ethylene glycol^[14]), carbon-based (e.g. graphene, carbon nanotubes, carbon dot),^[15,16] aggregation-induced emission (AIE) dots^[17,18] and mesoporous (e.g. mesoporous silica, mesoporous glasses)^[19–21] nanomaterials have been explored. Among the above, materials possessing fluorescence

as a diagnosis probe is popular since fluorescence imaging provides high sensitivity and can also be used for molecular imaging.^[22] In the above, dyes have been largely exploited as a fluorescent probe since they have high quantum yield, enhanced absorption coefficient, easy conjugation and familiarity with the biologist. However, fluorescent dyes possess many lacunae such as rapid photobleaching, narrow stokes emission, lesser excited state lifetime, overlapping with background emission, etc. These challenges can be overcome by lanthanide ion-doped nanoparticles which possess large stokes emission, enhanced lifetime, photostability and rich optical and magnetic properties due to their highly shielded 4f-electrons.^[23] We note that many lanthanide-doped nanomaterials have been largely exploited as bioimaging agent but not as theranostic agents.^[24,25] In this regard, few lanthanide-based nanoparticles are explored as theranostic vehicles, for example, core-shell structured NaLuF₄nanorods@polydopamine,^[5] Gd/Eu-doped ZnO NPs,^[6] lanthanum oxyiodide (LaOI) nanosheets,^[8] polyacrylic acid (PAA)-functionalized porous BiF₃:Yb,Er,^[9] NaLnF₄@MOF–Ln nanocomposites,^[10] zwitterionic gadolinium (III) (Gd(III))-complexed dendrimer-entrapped gold nanoparticles,^[11] etc. Though the above reports show promising results, most of the studies are in the preliminary level and they face the following challenges: (1) do not possess targeting ability, (2) limited drug loading, (3) extensive surface chemistry to conjugate the drug, which can affect the photoluminescence properties, (4) lack of information regarding *in vitro*, immunocompatibility and *in vivo* toxicity, (5) lack of information regarding bio-distribution and targeting efficiency and (6) use of high concentration for *in vivo* imaging. Further, most of the theranostic materials do not reach the clinical trials since they lack either one or combination of the following qualities such

[a] C. Rajalekshmi Dhanya,^{+, #} S. Sainulabdeen, M. S. Soumya, A. Abraham
Department of Biochemistry
University of Kerala
Kariavattom, Thiruvananthapuram 695581, Kerala (India)
E-mail: annieab2013@gmail.com

[b] J. Jeyaraman,⁺ Dr. S. Sivakumar
Department of Chemical Engineering
Center for Nanoscience and Center for Environmental Science and Engineering
Institute of Technology Kanpur
Kanpur, Uttar Pradesh (India)
E-mail: srisiva@iitk.ac.in

[†] These authors contributed equally to this work.

[‡] Presently affiliated to Government College, Kariavattom, Thiruvananthapuram, Kerala, India

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cptc.202000189>

An invited contribution to a Special Collection on Bioimaging

as stealth property, immuno-compatibility and biocompatibility at both in vitro and in vivo levels. Thus, it is essential to investigate the imaging and therapeutic capability, toxicity profile, biodistribution and clearance in the in vivo models. To this end, we report in vivo imaging, biodistribution and in vivo toxicity of Tb^{3+} -doped lanthanum vanadate ($LaVO_4$) nanoparticles encapsulated in doxorubicin-loaded PEGylated polymer capsules in mice models. We recently reported that lanthanide-doped nanoparticle loaded PEGylated polymer capsules possess good immuno-compatibility in dendritic cells/macrophages, cyto-compatibility, stealth property due to PEGylation, enhanced in vitro targeting ability.^[23,26]

Doxorubicin-loaded PEGylated polymer capsules encapsulating $LaVO_4:Tb^{3+}$ nanoparticles have been administered in tumor mice model (solid tumor formed by injecting Dalton's lymphoma ascites (DLA) cells). Bright green Tb^{3+} fluorescence has been observed from the tumor region at lower concentration ($\sim 3.4 \mu M$ $LaVO_4/kg$ body weight) of nanoparticles suggesting its enhanced targeting ability. Further, biodistribution studies confirm the presence of higher concentration of lanthanide ions in tumor compared to other tissues which further support its enhanced targeting ability. Moreover, there is a significant reduction in the tumor volume compared to control experiments (free doxorubicin (dox) treated and buffer treated) which supports its enhanced therapeutic ability. This was confirmed by a detailed study on tumor markers which indicates normal levels of enzyme activity in animals treated with dox loaded polymer capsules. Normal parameters in normal control mice treated with drug loaded polymer capsules indicate its biocompatibility. The condition/shape/color/weight of spleen did not show any difference between the capsules treated and untreated mice suggesting its immuno-compatibility. Our study has been summarized in Scheme 1.

2. Materials and Methods

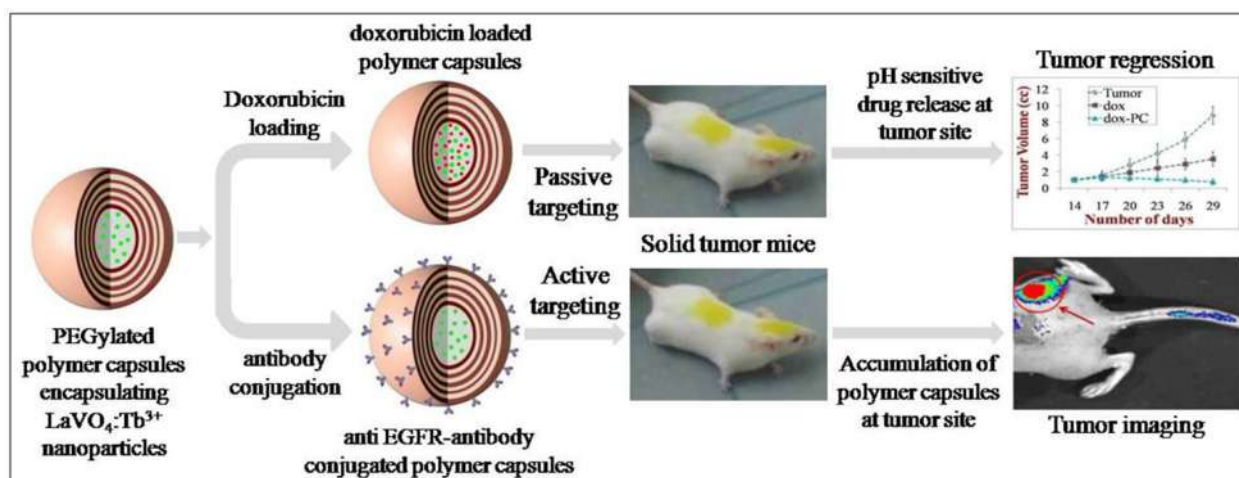
2.1. Materials

All chemicals, biochemicals and reagents were of analytical grade. Lanthanum nitrate hydrate ($LaNO_3 \cdot xH_2O$), terbium nitrate pentahydrate ($Tb(NO_3)_3 \cdot 5H_2O$), polyethylene glycol (PEG) (Mw. 10,000), poly (sodium 4-styrene-sulfonate) (PSS) Mw. 70,000, poly (allylamine hydrochloride) (PAH), Mw.56,000, bisamine (polyethylene glycol) Mw 6,000, N-Hydroxysuccinimide, (1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide) (EDC), N-hydroxysuccinimide (NHS) and doxorubicin hydrochloride (dox) were purchased from Sigma Aldrich. All the lanthanide salts were 99.9% pure and used without further purification. Polyethylene imine (PEI, Mw.70000 branched), Alfa Aesar; citric acid (Qualigens), liquor ammonia (about 24%, Merck), ammonium fluoride (Merck), sodium metavanadate Lobachemie; hydrofluoric acid (40% w/v) sd-fine chem., tetraethylorthosilicate (Fluka). Anti-Epidermal growth factor receptor (EGFR) antibody was obtained from invitrogen, India.

Enzyme based assays for detection of alkaline phosphatase (ALP) and lactate dehydrogenase (LDH) were carried out using ERBA biological diagnostic kit.

2.2. Synthesis of dox-Loaded PEGylated Polymer Capsules and Drug Release

PEGylated polymer capsules encapsulating $LaVO_4:Tb^{3+}$ nanoparticles were synthesized by a two-step process.^[23,26,27] Firstly, $LaVO_4:Tb^{3+}@silica$ core-shell particles were formed by sol-gel method. Secondly, Layer-by-Layer (LbL) on $LaVO_4:Tb^{3+}@silica$ particles was made by PEI as first layer, followed by consecutive assembly of 8 layers of PSS/PAH and PEG as ultimate layer. Finally, silica core is removed using HF to form polymer capsules encapsulating nanoparticles. The morphology of the polymer capsules and encapsulation of nanoparticles were



Scheme 1. Summary of overall study: Doxorubicin loaded PEGylated polymer capsules were passively targeted and studied for tumor regression, and anti-EGFR antibody conjugated PEGylated polymer capsules were actively targeted to tumors for imaging.

visualized from SUPRA[®] Series Ultra High-Resolution Field Emission-Scanning Electron Microscope FE-SEM (SEM) and transmission electron microscope (TEM) images obtained using FEI Technai Twin microscope with high contrast and resolution at 20 kV to 120 kV.

2.3. Drug Loading

40 µg of dox solution was added to the PEGylated polymer capsules and kept in an orbital shaker overnight, after which it was centrifuged. Drug loading was followed by measuring absorbance of supernatant at 480 nm, before and after encapsulation using JASCO-L-670 spectrophotometer. Difference in absorbance gave the quantity of drug encapsulated.^[4,28]

Entrapment efficiency was calculated as follows [Eq. (1)]:

$$\text{Entrapment efficiency} = \frac{\text{Amount of drug incorporated} \times 100}{\text{Amount of the drug initially taken}} \quad (1)$$

The encapsulation of drug was also confirmed by conducting FTIR analysis using PerkinElmer Spectrum 100 FTIR Spectrometer. Further, absorption spectra and emission spectra of dox loaded polymer capsules were also measured (using JASCO-L-670 spectrophotometer and Horiba Scientific FluoroMax-4 Spectrofluorometer respectively).

2.4. Drug Release

The release profile of doxorubicin (dox) from capsules at 37 °C was determined using Horiba Scientific FluoroMax-4 Spectrofluorometer, at acidic pH (pH 5.8) and physiological pH (pH 7.4). dox-loaded PEGylated polymer capsules were dispersed in 7 mL buffered saline of corresponding pH and aliquoted into seven equal fractions. Each fraction was centrifuged at various time periods and fluorescence of supernatant was measured ($\lambda_{\text{ex}} = 480 \text{ nm}$; $\lambda_{\text{em}} = 560 \text{ nm}$).^[4,29]

2.5. In Vivo Studies

Female Swiss albino mice (2-3 months) weighing 20–25 g were reared in the department animal house and fed with standard laboratory pellet diet. They were housed in an environment of controlled temperature, humidity and light/dark cycle and drinking water was provided *ad libitum* and were weighed weekly. All ethical guidelines were followed for carrying out animal experiments in firm compliance with Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA) Government of India and ARRIVE guidelines. The experiments have been approved by Institutional Animal Ethical Committee (IAEC) as per sanction order IAEC-KU-5/2010-'11-BCAA. Animal behavior, food and water consumption and survival were observed during the experimental period.

2.5.1. Solid Tumor Induction

Mature DLA cells, maintained in the peritoneal cavity of mice, were aspirated and suspended in Phosphate buffered saline (PBS). Solid tumor was induced by subcutaneous injection of DLA cells (1×10^6 cells in 0.1 ml PBS /animal) to the hind limb of experimental mice.^[30]

2.5.2. Biodistribution of PEGylated Polymer Capsules by Elemental Analysis of Lanthanide by ICP-MS

Solid-tumor-bearing mice were administered with PEGylated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles at a dose of 3×10^9 capsules/kg body weight. The mice were sacrificed 24 h post injection. Organs including liver, kidney, spleen, heart, lungs and tumor were excised and blood was drawn. Tissues were thoroughly washed with PBS, blotted with filter paper and weighed. The whole organ samples and blood were digested with 70% nitric acid at a temperature of 110–120 °C for 4–5 h. The remaining undissolved solids were removed by filtering through a membrane and the clear solution obtained was diluted with distilled water for ICP-MS analysis.^[27]

2.5.3. Tumor Regression Studies

The animals were randomized into six groups: Group I and II included normal mice and group III to VI comprised of tumor bearing mice. Animals in different groups were identified by distinctly staining their fur yellow using picric acid (1.2% W/V). A schematic representation of experimental detail is given in Figure 3.

The experiment was performed when the tumor size reached $\approx 1.0 \text{ cc}$ after two weeks of administration of DLA cells. Group I comprising normal mice and group III solid-tumor-bearing mice, were given PBS. Group IV & VI were administered with dox loaded PEGylated polymer capsules encapsulating nanoparticles (capsules corresponding to 100 µg dox/kg body weight) once every three days from day 14 to day 30 (14th, 17th, 20th, 23rd, 26th, 29th day). Group V animals were injected with 100 µg/kg body weight of free dox. Equivalent number of dox-loaded capsules were injected to group II mice at same regularity. The tumor radii were measured every third day and the tumor volume was estimated using the formula, $V = 4/3\pi r_1^2 r_2$, where r_1 and r_2 are the radii of tumors at two different directions.^[31]

Administration of PBS/doxorubicin/doxorubicin loaded polymer capsules was stopped on day 30. At the end of experimental period (day 30), animals from group I to V were sacrificed. Blood was collected and serum was separated for assay of lactate dehydrogenase (LDH) and alkaline phosphatase (ALP). Liver and kidney tissues were taken for determining activity of β -D glucuronidase and myeloperoxidase (MPO).^[32] Comet assay was conducted in tumor tissues to study DNA

level changes.^[33] Histology was performed in tumor and liver tissues. Spleen size and weight was monitored.

Animals in Group IV and VI, that were administered with dox loaded polymer capsules were sacrificed at different time intervals and various parameters related to tumor regression (including spleen size) were studied and compared. Animals in group VI were maintained till day 90 without administration of drug loaded polymer capsules to check if tumor recurred during this period. On day 90, they were sacrificed and all the above parameters were analyzed.

All the statistical calculations were carried out using SPSS software program (version 17 for Windows). The values were expressed as mean \pm standard deviation. The data were analyzed using one-way ANOVA and significant difference of means between groups was determined using Duncan's multiple range tests at the level of $p < 0.05$. Same alphabet on error bars indicates no significant change between groups, whereas different alphabets indicate significant difference of values between groups at $p < 0.05$.

2.5.4. Tumor Imaging Studies

The PEGylated polymer capsules were conjugated to anti-EGFR antibody and this was used for in vivo imaging in solid tumor bearing mice.

2.5.4.1. Antibody Conjugation

Surface modification of the polymer capsules was carried out by using EDC-NHS coupling.^[34] PEGylated polymer capsules encapsulating nanoparticles were suspended in MES (4-morpholino-ethane-sulfonic acid) buffer (1 mL, 500 mM, pH 6.1) followed by addition of NHS (2.3 mL, 50 mg/mL) and EDC (1.2 mL, 10 mg/mL) at room temperature and incubated for 30 min. This reaction mixture was centrifuged and washed in excess MES buffer (50 mM). Polymer capsules were resuspended in MES buffer (9 mL, 50 mM) containing anti-EGFR antibody (10 μ g/mL, pH 6.1) and put in orbital shaker at 1500 rpm for 1 h. Finally, the antibody modified capsules were centrifuged and washed thrice with MES buffer (50 mM).^[23,34] To confirm antibody conjugation, the anti-EGFR conjugated PEGylated polymer capsules were incubated with FITC (Fluorescein isothiocyanate)-linked secondary antibody, washed with MES buffer and viewed under fluorescent microscope (Olympus CKX-41, Japan).

2.5.4.2. In Vivo Fluorescence Imaging

The mice were subjected to in vivo imaging studies after 2 to 3 weeks when the tumor volume reached ≈ 1.0 cc. For imaging experiments, the animals were randomized into two groups: Group I and group II. Animals in the two groups were identified by staining their fur yellow using picric acid (1.2% W/V). Tumor control mice in group I received only PBS. Tumor mice in group

II was injected with anti-EGFR antibody conjugated PEGylated polymer capsules encapsulating nanoparticles (3×10^9 capsules/kg body wt.) intravenously through tail vein. After 3 h, mice were anesthetized by intra-peritoneal injection of ketamine 80 mg/kg body weight and then subjected to in vivo imaging (Xenogen's IVIS[®] 200 Series) with a 480 nm wavelength excitation and the fluorescence images were recorded.

3. Results and Discussion

Synthesis of PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles have been performed in three major steps: (a) Sol-gel synthesis of LaVO₄:Tb³⁺@silica, (b) formation of layer-by-layer assembly, and (c) removal of silica core to yield nanoparticles encapsulated capsule. PEGylated polymer capsules encapsulating lanthanide doped nanoparticles were characterized as in the previously reported study.^[23,27] Figure 1A shows TEM image of ~ 500 nm sized PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles in which contrast difference suggests entrapment of nanoparticles inside capsules. High resolution Transmission electron microscope (HRTEM) image of inset shows lattice fringes suggesting crystalline nature of nanoparticles. The samples are dried for acquiring TEM images and aggregation observed is due to drying effect. SEM images in Figure 1B shows morphology of capsules and absence of nanoparticles on its surface. Further, emission spectra of PEGylated polymer capsules encapsulating

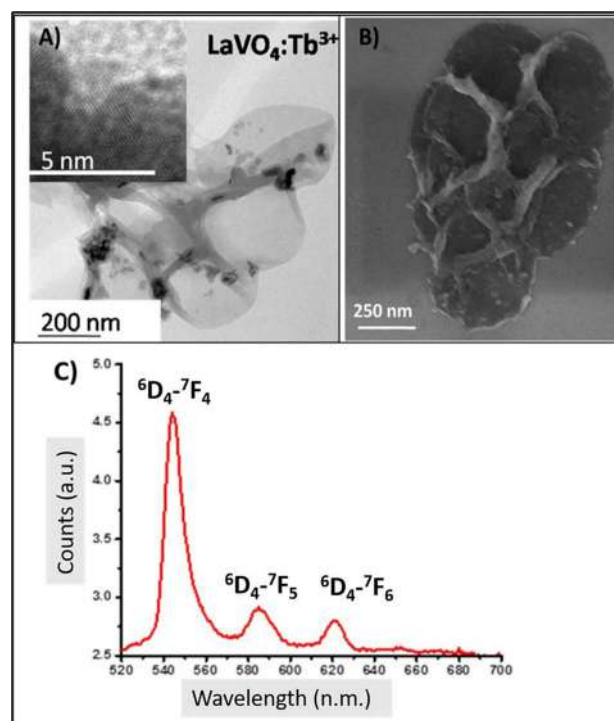


Figure 1. PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles. (A) TEM (Inset: HR-TEM), (B) SEM, and (C) Emission spectrum of Tb³⁺ ions of PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles (Excitation wavelength = 488 nm, high bandpass filter used = 515 nm).

LaVO₄:Tb³⁺ nanoparticles with excitation at 488 nm (direct excitation of terbium ions) shows emission peaks at 545 nm, 585 nm and 620 nm (Figure 1C).

3.1. Loading of Doxorubicin Inside Capsules

Anthracyclin drug doxorubicin was loaded into the PEGylated polymer capsules as mentioned in the experimental section. Entrapment efficiency was found to be 50% (Figure S1). Loading of doxorubicin was further confirmed by FTIR analysis. FTIR spectrum of free doxorubicin shows peaks at 1430 cm⁻¹ due to aromatic C=C and at 1673 cm⁻¹ due to carbonyl group. More noteworthy is the broad and strong stretching bands due to O–H stretching at 3437 cm⁻¹. As seen in Figure S2, dox has several –OH groups in it and O–H stretching is characterized by peak within the range 3200–3600 cm⁻¹. Polymer capsules (PC) and doxorubicin loaded polymer capsules (dox-PC) are showing comparable peaks. But the latter exhibits a strong broad peak at 3481 cm⁻¹ which is absent in polymer capsules (PC) alone, and present in free dox pointing towards dox loading. In addition, UV-Visible and fluorescence spectroscopy was carried out for polymer capsules before and after loading of doxorubicin. Broad peak with maximum absorbance at 480 nm has been observed in the polymer capsules loaded with doxorubicin which is attributed to doxorubicin whereas the bare capsules show weak absorption which is attributed to polymer layers (Figure S3). It is noted that the absorption corresponding to Tb³⁺ is not visible due to the 4f–4f forbidden transition. Further, emission spectrum shows broad emission (585 nm) corresponding to dox is observed in the case of polymer capsules loaded with doxorubicin which proves the loading of dox (Figure S4).

3.2. Drug Release Study

In order to study the release of drug at varying pH, the capsules loaded with doxorubicin were subjected to acidic pH (5.8) and physiological pH (7.4). PAH is a weak polyelectrolyte with pKa value 8.7 in a salt free solution, and after complexing with PSS the pKa value is changed to 10.7.^[35] From the release profile (Figure S5) it is clearly evident that the drug release rate is higher in the case of acidic pH compared to physiological pH as the polyelectrolytes become uncharged in response to pH. After 72 h, ~30% drug was released at physiological pH, whereas at acidic pH, ~90% dox was released (Figure S5). It is desirable as tumor micro-environment is acidic when compared to that in normal tissues. Further, it was noted that there is no initial burst effect at both pH which can be due to the LbL assembly. There are reports suggesting that undesirable burst release can be controlled by layer-by-layer assembly.^[36] In the present study, we observed a gradual increase in rate of dox release at both pH which is desirable since it can prevent the detrimental drug leakage during the circulation which can potentially avoid the toxic side effect and reduced drug potency at the tumor in vivo. It is illustrious that the burst

release at the physiological pH may induce toxicity to the other organs.^[37] We note that influence of pH on porosity and morphology of PSS/PAH capsules has been previously reported and the current results match with them.^[38–40] Lower the pH, more the dox released from polymer capsules. At lower pH, swelling of PAH/PSS films happen due to variation in the charge density of polyelectrolyte chains which induce the electrostatic repulsion between PAH/PSS multilayers and thus facilitating drug release (Scheme S1). The mechanism of release has been studied in the previous reports.^[41,42] As the drug release from polymer capsules is higher at acidic pH, more availability of doxorubicin in tumor tissues can be achieved and this can be exploited to improve therapeutic efficacy in vivo.

3.3. In Vivo Studies

In vivo studies have been performed with the nanoparticles loaded polymer capsules in order to evaluate their applicability. We note the following for clarification: (1) Biodistribution studies were performed with PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles without antibody conjugation (passive targeting) in tumor model mice. (2) PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles (PC) with antibody conjugation (active targeting) have been used for imaging experiments. (3) PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles without antibody conjugation (passive targeting) have been used for drug delivery experiments.

3.3.1. Biodistribution

We have previously reported that maximum percentage of lanthanum is found in liver, 6 h post administration and that it is cleared through urine after 24 h in normal mice.^[27] In the present study we followed the biodistribution of PEGylated polymer capsules encapsulating LaVO₄:Tb³⁺ nanoparticles (PC) in tumor mice. Solid tumor was induced in the hind limb of Swiss Albino mice by injecting Dalton's Lymphoma Ascites (DLA).^[30,43] Biodistribution of PEGylated polymer capsules at 24 h post-injection in solid-tumor-bearing mice was determined by quantifying the amount of lanthanum in tissue extracts by Inductively Coupled Plasma Mass Spectrometry (ICP-MS). Highest percentage (~50%) of polymer capsules equivalent to concentration of lanthanum (in parts per billion (ppb)) was detected in tumor tissue supporting its size dependent passive targeting ability. This may be attributed to extravasations of the drug loaded polymer capsules to the tumor tissue and its retention by virtue of EPR (enhanced permeability and retention) effect. The hyper permeable nature of the tumor microcirculation is reported in various reports and it is noted that tumors grown subcutaneously exhibits a characteristic pore cut-off size ranging from 200 nm to 1.2 μm.^[44–46] Our results show ~50% of capsules in tumor site and lesser concentration in other tissues particularly kidney and spleen (Figure 2).

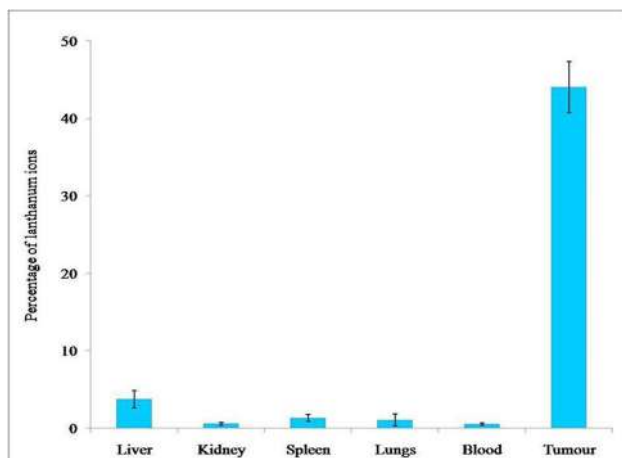


Figure 2. Graph representing ICP-MS data of lanthanum ion (in percentage) in various tissues of tumor bearing Swiss Albino mice 24 h after administration of PEGylated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles (3×10^9 capsules/kg body weight). Highest percentage of lanthanum ions is found in tumor tissue proving the efficiency of passive targeted PEGylated polymer capsules.

3.3.2. Tumor Regression Study

The therapeutic potential of dox-loaded PEGylated polymer capsules (PC) was determined by tumor regression studies in DLA induced solid tumor bearing Swiss Albino mice. Tumor regression studies indicate reduction in tumor volume after treating the mice with dox-loaded capsules. Figure 4A to 4F shows panel of mice bearing tumor cells treated with dox-loaded capsules and free-dox. It has been observed that tumor size is reduced in mice injected with dox-loaded capsules more

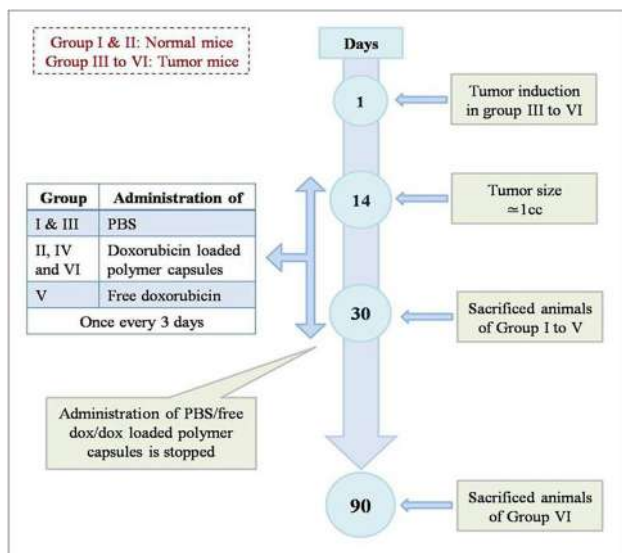


Figure 3. Experimental design showing animal treatment procedure: Group I = normal mice receiving PBS; Group II = normal mice administered with dox-PC; Group III = tumor control receiving PBS; Groups IV and VI = tumor mice administered with dox-PC; Group V = tumor mice administered with dox. Animals in groups I–V are sacrificed on day 30 where as those in group VI are sacrificed on day 90.

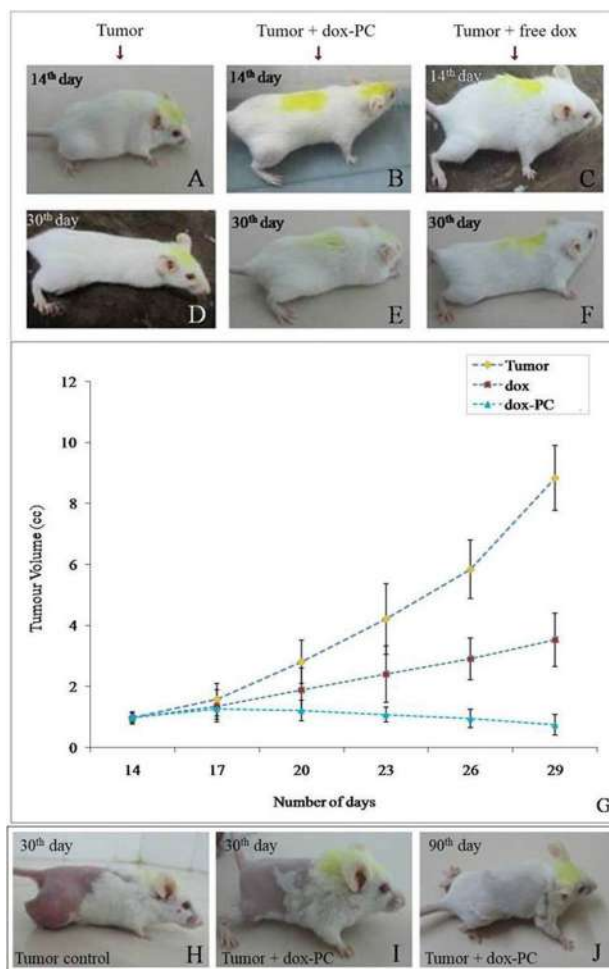


Figure 4. Tumor regression in mice. A, D: Tumor control (group III); B, E: dox loaded capsules treated (group IV); C, F: Free dox treated (group V). Tumor regression in mice treated with dox loaded capsules is more prominent when compared to free dox treated animals. G shows the tumor volume with respect to time which reduced significantly in mice of groups IV and V. Decrease in tumor volume is more pronounced in mice treated with dox loaded capsules than that in free dox treated animals, when compared with tumor control. J indicates that tumor has disappeared and not recurred after 90 days in animals treated with dox loaded capsules.

prominently than in mice injected with free-dox. From Figure 4G, it is confirmed that tumor volume is reduced to less than 2 cm^3 in case of dox-loaded capsules in contrast to free-dox which is slightly higher. Figure 4J clearly demonstrates that tumor has not recurred in group VI mice even after 90 days.

In order to further study the effect of dox-loaded capsules in tumor mice, we performed quantification of tumor marker enzymes in liver and kidney, serum enzymes, genotoxicity assay, histology examinations (tumor and liver) and spleen size/weight measurements.

3.3.2.1. Marker Enzymes

Therapeutic efficacy of dox loaded polymer capsules was further established by comparing the activities of β -D-glucuronidases, myeloperoxidase, LDH and ALP in treated and

untreated tumor bearing animals. β -D-glucuronidases are members of the glycosidase family of enzymes that catalyze hydrolysis of β -D-glucuronic acid residues from the non-reducing end of mucopolysaccharides, also referred to as glycosaminoglycans such as heparin sulfate. This facilitates degradation of extra cellular matrix (ECM) and release of heparin sulfate bound biomolecules including cytokines. The remodeling of ECM is important for various physiological and pathological processes including inflammation, tumor angiogenesis and metastasis.^[47–49] Activity of this enzyme in serum and tissues have been reported to be elevated in inflammation and tumor condition.^[48,50–52] Similar observation has been found in liver and kidney of solid tumor bearing mice.^[52]

Myeloperoxidase (MPO) is a heme containing enzyme stored in the azurophilic granules of polymorphonuclear neutrophils and macrophages. Measurement of MPO activity is frequently used as a marker of neutrophil infiltration into tissues.^[53,54] Cancer cells exhibit an enhanced rate of glycolysis to produce adequate energy to meet demands of rapidly proliferating tumor cells, even under normal oxygen concentrations, a phenomenon known as Warburg effect.^[55] Under hypoxic condition in cancer, NAD^+ is regenerated from NADH by the reduction of pyruvate to lactate by LDH, in order to ensure continuation of glycolysis. Serum LDH is a key biomarker in many types of cancer. LDH and lactate production are involved in tumor initiation, invasive potential, metastasis and recurrence.^[56] It is a non-specific tumor marker and raised levels of LDH signify rapid cell turnover rate and large tumor burden.^[57–59]

ALP is a group of enzymes that catalyzes phosphate ester hydrolysis in an alkaline environment, generating an organic radical and inorganic phosphate.^[60] Serum ALP levels are frequently elevated in patients with metastatic colorectal cancer and liver disease^[61] and DLA bearing mice.^[62,63] Activities of β -D-glucuronidase and MPO (Figure 5) in tissues and LDH and ALP (Figure 6) in serum were found to be elevated in tumor bearing mice (group III). This is significantly lowered in dox loaded polymer capsules treated mice (group IV) as well as in free dox treated mice (group V), with the former being more prominent. The values reverted to near normal in dox loaded polymer capsules treated mice sacrificed after 90 d (group VI). The enzyme activities in group II mice were found to be similar to that in group I indicating non-toxicity of dox loaded polymer capsules in normal mice. Comparable LDH activity in dox loaded polymer capsules administered normal animals indicates that it does not produce any toxicity.

3.3.2.2. Comet Assay

Genotoxicity measurements were carried out by performing single cell gel electrophoresis assay and the comets were quantified using TriTekCometScoreTM v1.5 comet scoring software. DNA damage in tumor tissue of animals in group IV and V is indicated by the extent of comet tailing (figure 7A to 7C). When compared to free dox treated mice, the degree of DNA damage is higher in tumor tissue of animals treated with

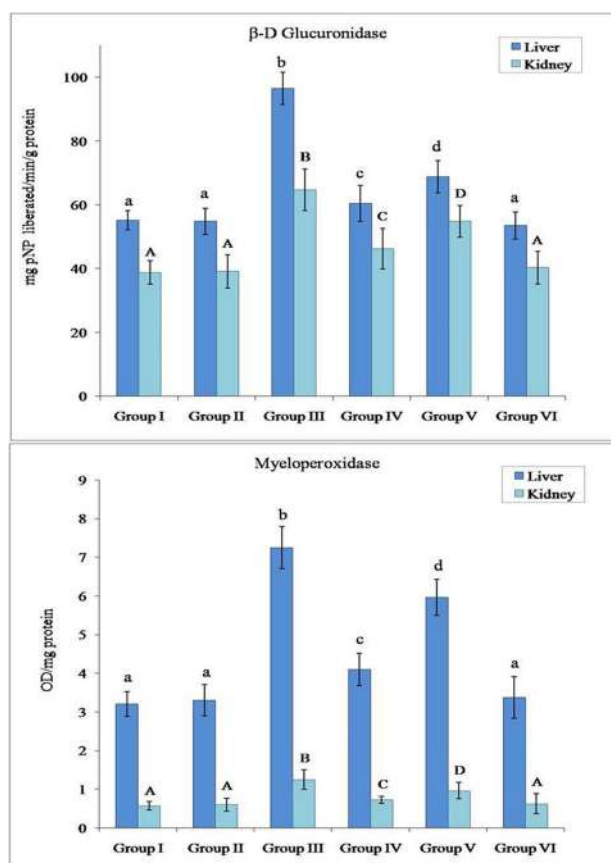


Figure 5. Activity of β -D-glucuronidase and Myeloperoxidase in liver and kidney of mice, Group I: Normal control; group II: Normal control mice treated with dox loaded capsules; group III: Solid tumor control; group IV: Solid-tumor-bearing mice treated with dox loaded capsules; group V: Solid-tumor-bearing mice treated with free dox; group VI: tumor mice treated with dox loaded capsules; sacrificed after 90 days.

dox loaded polymer capsules, signified by the higher comet parameters (Figure 7D to 7F) like tail length, percentage DNA in tail, tail moment and Olive moment (Figure 7G to 7J). Comet tail moment corresponds to the extent of DNA damage in individual cells. Tail moment, which is the product of the tail length and the fraction of total DNA in the tail, is calculated automatically by the software as an average for 50 cells selected for measurement. Olive tail moment is also obtained by the software for each cell analyzed. This parameter represents the product of the percentage of total DNA in the tail and the distance between the centers of the mass of head and tail regions.^[64] It has long been reported that dox induces p53 mediated tumor cell apoptosis and thus DNA damage.^[65,66]

3.3.2.3. Histological Studies

Histological examinations are carried out in tumor and liver tissues to study the changes before and after treating with dox-loaded capsules. Tumor histology of untreated solid tumor, as viewed under light microscope, showed aggregation of tumor cells spread subcutaneously (Figure 8A). There are previous

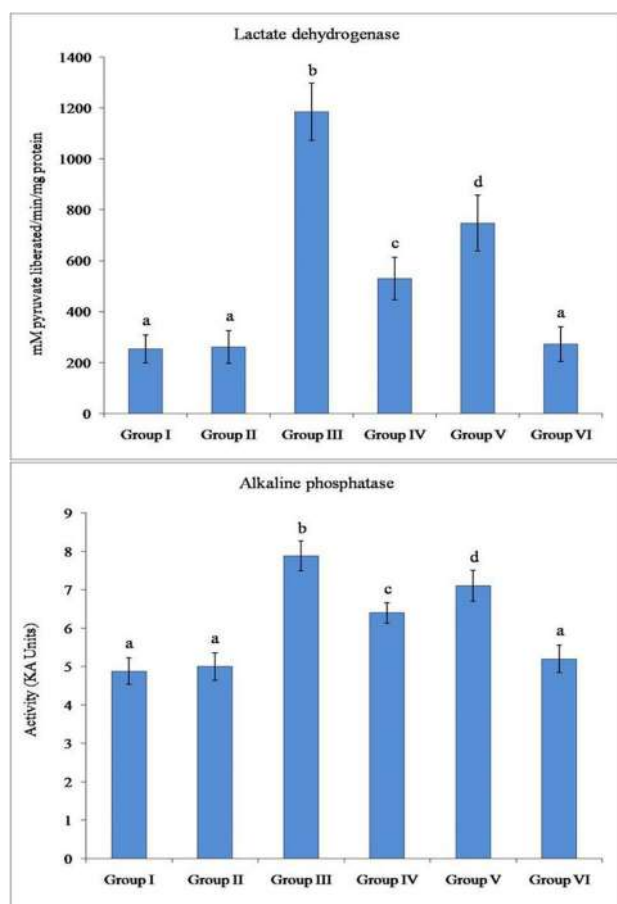


Figure 6. Activity of LDH and ALP in serum: Group I: Normal control; group II: Normal control mice treated with dox loaded capsules; group III: Solid tumor control; group IV: Solid-tumor-bearing mice treated with dox loaded capsules; group V: Solid-tumor-bearing mice treated with free dox; group VI: tumor mice treated with dox loaded capsules; sacrificed after 90 days.

reports describing similar observations.^[43,67,68] Sections of tumor from mice treated with free dox (Figure 8C) as well as that with dox loaded polymer capsules (Figure 8B) illustrated tumor destruction. Dead cells observed are indicated by arrows in the figures. The number of tumor cells was significantly reduced in dox loaded polymer capsules treated group when compared to free dox treated animals.

Liver sections of mice bearing solid tumor showed accumulation of metastatic tumor cells (blue arrow) around portal blood vessels with hemolysed RBCs (red arrow) (Figure 8F). Reports are available on analogous results in tumor bearing animals.^[58,63] The amount of metastatic cells has come down to a large extent and hemolysed RBCs disappeared in animals treated with dox loaded polymer capsules (Figure 8G). The decline in tumor cells is comparatively lower in the liver sections of those treated with free dox (Figure 8H). Normal histology of liver was almost reinstated in dox loaded polymer capsules treated tumor mice sacrificed after 90d (Figure 8I) and was comparable to that in normal control (Figure 8D). Normal liver histology of group II animals treated with dox loaded polymer capsules pointed towards non-toxicity of the drug loaded capsules (Figure 8E).

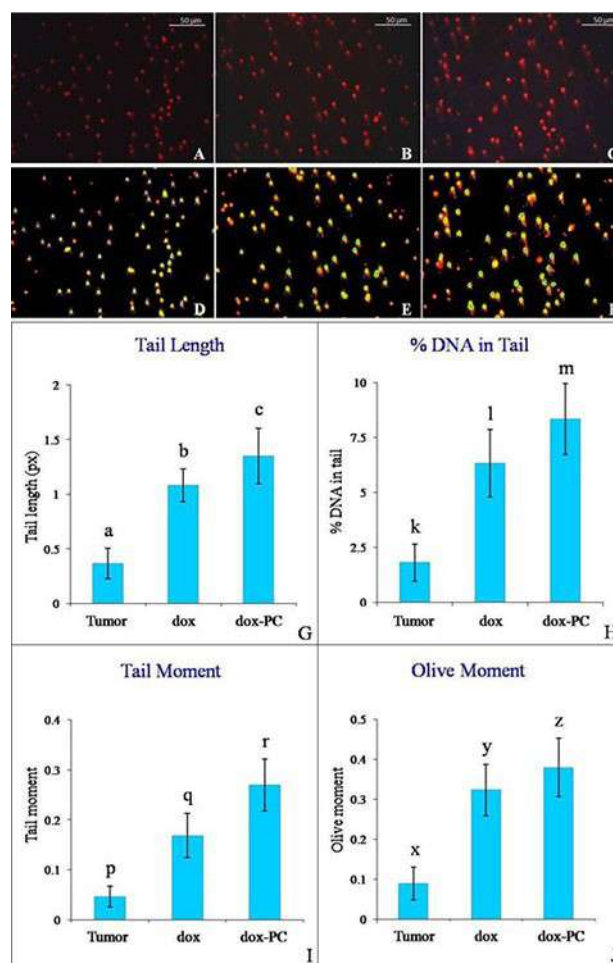


Figure 7. Comet assay for genotoxicity assessment in tumor tissue. A–C: Fluorescence micrographs; D–F: Respective images of comet scoring. A, D: Solid tumor control; B, E: Solid-tumor-bearing mice treated with free dox. C, F: Solid-tumor-bearing mice treated with dox loaded capsules. Graphs G, H, I and J represents corresponding comet scores.

3.3.2.4. Spleen Size

Another clinical manifestation associated with malignancy is enlarged spleen as a result of immunological activity against the tumor. Splenic enlargement is owing to immunological reaction evoked by antigen liberated from the tumor.^[69–72] Distant tumors are shown to impact splenic function. During tumor progression, monocytes formed by spleen can migrate to tumor microenvironment and create tumor-associated macrophages and tumor bearing mice have an increase in monocytic cells in the spleen.^[73] It is also proposed that metastasis to the spleen is one of the causes for cancer associated splenomegaly. In the present study, enlarged spleen was observed in solid tumor bearing mice. Spleen size in mice treated with dox loaded polymer capsules and those with free dox were not as large as in tumor mice. It is perhaps because tumor had been obliterated sooner than the spleen was exposed to adequate infiltration with cancer cells and antigenic stimulation. The size of spleen was reverted to normal in dox loaded polymer capsules treated group VI tumor mice

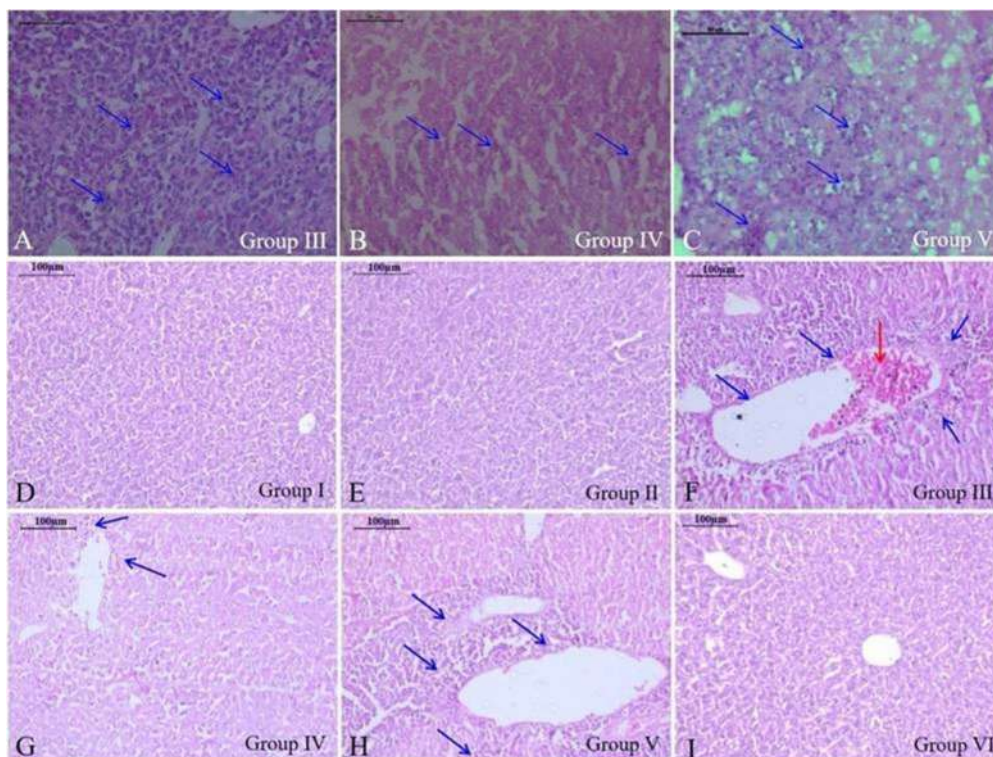


Figure 8. A–C: Tumor histology. A: Group III (solid tumor control) showing aggregation of tumor tissue cells. B: Group IV (tumor bearing mice treated with dox loaded capsules). C: Group V (tumor bearing mice treated with free dox). Dead cells are indicated by arrows. A lesser number of tumor cells is observed in (B) when compared to (C). D–I: Liver histology. D: Normal control, E: Normal control mice treated with dox loaded capsules, F: Solid tumor control, G: Tumor mice treated with dox loaded capsules. H: Tumor mice treated with free dox. I: Tumor mice treated with dox loaded capsules, sacrificed after 90 days. Blue arrows indicate accumulation of metastatic tumor cells around portal blood vessels and red arrow shows hemolysed RBCs.

sacrificed after 90d. Values of spleen weight correlated with this observation. Spleen size and weight in group II animals was comparable to that in normal control indicating non-toxicity of dox loaded polymer capsules in normal mice (Figure 9).

3.3.3. Antibody Conjugation and In Vivo Imaging

A technique employed in bioimaging is to functionalize the surface of nanoparticles with antibodies so as to recognize specific domains or species like tumor markers.^[74] In addition to passive targeting to the tumors, active targeting involving ligand-conjugated capsules that target specific receptors on the cancer cells can be investigated as the second approach.^[75] In this study, we had verified if the PEGylated polymer capsules can be actively targeted to tumor for in vivo imaging by conjugating it to anti-EGFR antibody. In our study, anti-EGFR antibody was successfully conjugated to PEGylated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles. This was confirmed by conjugating FITC linked secondary antibody to anti-EGFR antibody conjugated PEGylated polymer capsules. The intensity of green fluorescence native to the polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles was enhanced due to conjugation of FITC linked secondary antibody as indicated in Figure S6 when compared to anti-EGFR antibody

conjugated polymer capsules. Solid-tumor-bearing mice were subjected to in vivo tumor imaging after intravenous administration of anti-EGFR antibody conjugated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles. Fluorescence signal was obtained at the tumor site 3 h after injection and the tumor was distinguishable with good fluorescence contrast indicating high tumor uptake of anti-EGFR antibody conjugated polymer capsules (Figure 10). EGFR is over-expressed on malignant cell surfaces allowing capsules coated with anti-EGFR antibodies to bind preferentially over non-cancer cells.^[76–79] Further, the enhanced permeation and retention (EPR) effect due to size and leaky nature of tumor favors the accumulation of capsules at the tumor site which is evident from the biodistribution studies with polymer capsules without the antibody conjugation. Thus, we believe that the highly expressed EGFR and the EPR effect can cumulatively favor enhanced accumulation of anti-EGFR antibody conjugated PEGylated polymer capsules in the tumor. The control experiment that has been simultaneously done by imaging tumor bearing mice administered with PBS (Figure 10) does not show any signal at tumor site.

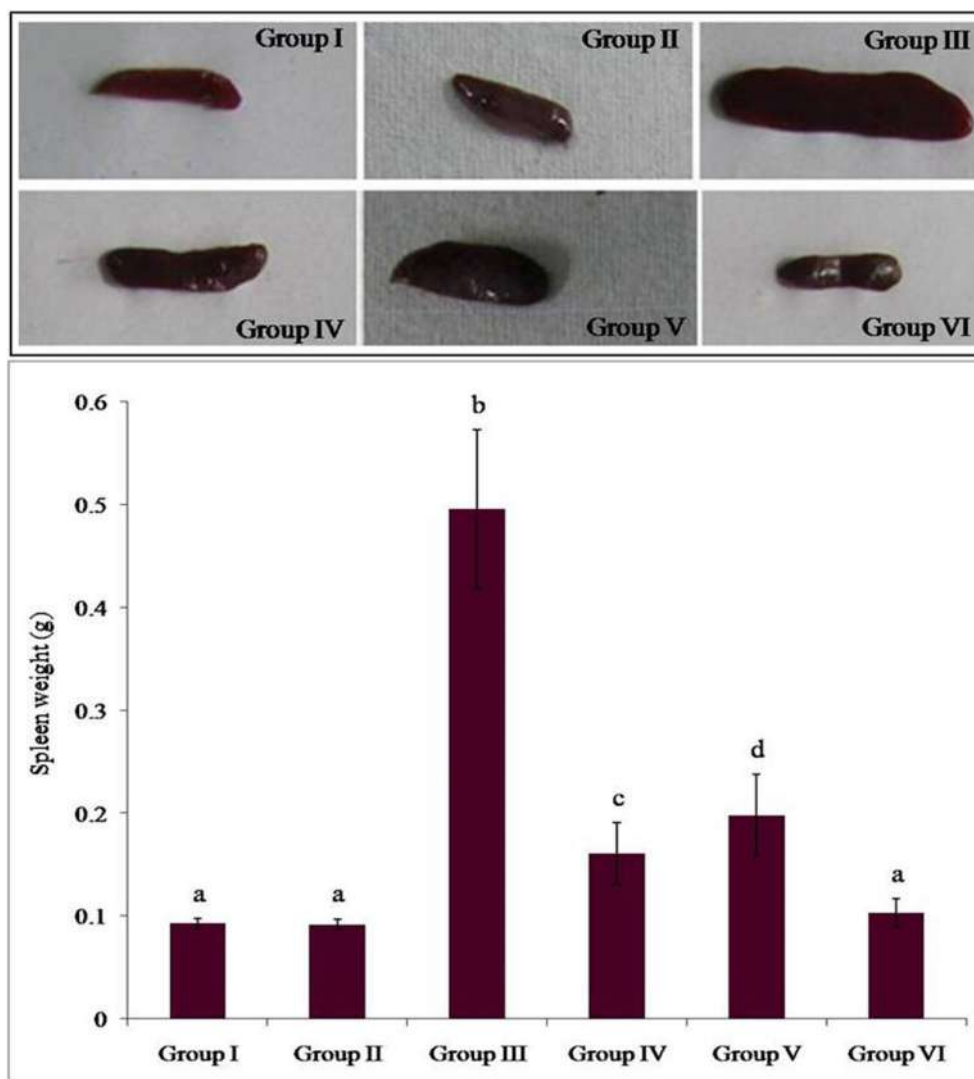


Figure 9. Photographs representing spleen size. Group I: Normal control; group II: Normal control mice treated with dox loaded capsules; group III: solid tumor control; group IV: Solid-tumor-bearing mice treated with dox loaded capsules; group V: Solid-tumor-bearing mice treated with free dox. Animals in groups I–V were sacrificed on day 30; group VI: tumor-bearing mice treated with dox loaded capsules; sacrificed on day 90. Graph represents corresponding spleen weight.

4. Conclusions

PEGylated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles were successfully employed for both passive and active targeting to solid tumor. Highest concentration of PEGylated polymer capsules was seen in tumor, pointing towards the size dependent passive targeting ability of the capsules. Biodistribution studies demonstrated lesser concentration in other tissues including kidney, liver and spleen. In vivo studies carried out in solid tumor bearing Swiss Albino mice showed reduction in tumor volume on treatment with dox-loaded PEGylated polymer capsules. It also showed reduction in the activity of tumor marker enzymes and levels of serum enzymes indicating tumor destruction due to action of doxorubicin. The genotoxicity assay (comet assay) showed Comets after treating with dox-loaded capsules which display DNA distortion induced by doxorubicin. Further, histological

studies also indicate the reduction in tumor cells in both tumor site and liver of mice treated with dox-loaded capsules. In addition, size and weight of spleen in dox-loaded PEGylated polymer capsules treated mice were also comparable to normal mice which are major indications that tumor cells are getting reduced after treating with dox-loaded PEGylated polymer capsules. It was also observed that, tumor did not reappear even after 60 days of discontinuing the administration of dox loaded polymer capsules. Anti-EGFR antibody conjugated PEGylated polymer capsules was found to be effective for in vivo fluorescence imaging in Swiss Albino mice bearing DLA tumor cells. Bright green Tb^{3+} fluorescence from the tumor site of tumor mice administered with anti-EGFR conjugated PEGylated polymer capsules suggested its enhanced active targeting ability. Therefore, PEGylated polymer capsules encapsulating $\text{LaVO}_4:\text{Tb}^{3+}$ nanoparticles are fabricated as potential theranostic vehicle for fluorescence imaging and drug delivery.

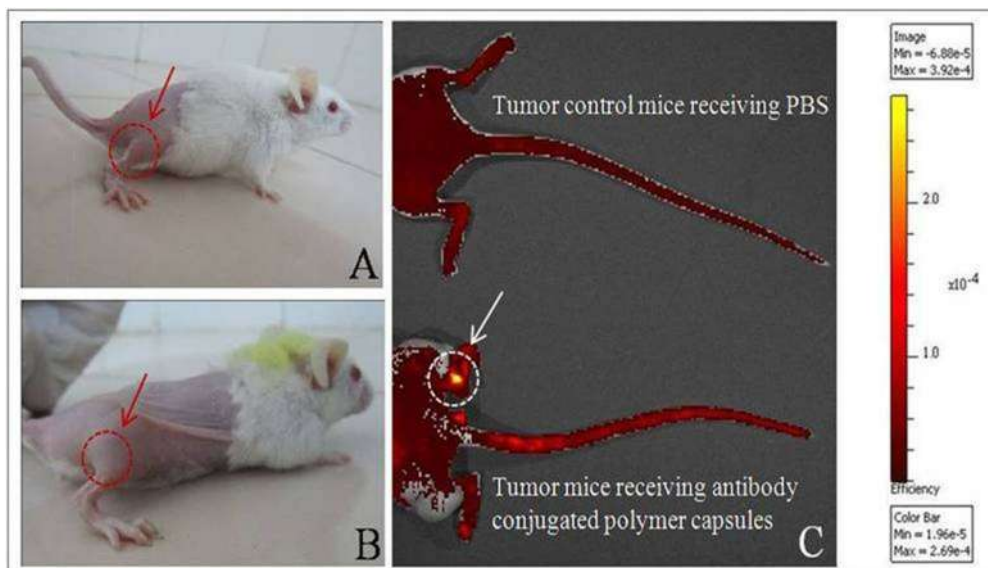


Figure 10. A: Photograph of control tumor bearing mice receiving PBS (group I), B: Photograph of tumor bearing mice administered with anti-EGFR-conjugated polymer capsules (group II), C: In vivo imaging 3 h after administration of PBS and anti-EGFR-conjugated polymer capsules to animals of groups I and II, respectively. The red arrows in (A) and (B) illustrate solid tumor in mice. The white arrow in (C) shows a signal at the tumor site of group II mice corresponding to emission of Tb³⁺ ions. No signal is obtained in group I control tumor mice receiving PBS. Color bar in (C) indicates the intensity of fluorescence.

Acknowledgements

We greatly acknowledge the grants from Nanomission, Department of Science and Technology (DST) (Grant numbers SR/NM/NS-1111/2013(G), SR/NM/NS-82/2016 (G)) and Department of Biotechnology (DBT) (Grant numbers: BT/PR13191/NNT/28/465/200, BT/PR21693/NNT/28/1180/20), India. We also acknowledge Dr. Lakshmi V Nair and Dr. R. S. Jayasree, Division of Biophotonics and Imaging, Sree Chitra Tirunal Institute for Medical Sciences and Technology, BMT Wing, Poojappura, Trivandrum, India for providing animal imaging facility. We furthermore express appreciation to Dr. AnanthaLekshmi, Veterinary Doctor, Department of Biochemistry, University of Kerala, Kariavattom, India and Dr. Sreeja S., Post-doctoral fellow, Cochin University of Science and Technology, Kerala, India for assisting in animal experiments.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Bioimaging · in vivo imaging · lanthanides · polymer capsules · theranostics

- [1] E. Porret, X. Le Guével, J. L. Coll, *J. Mater. Chem. B* **2020**.
- [2] B. Kłębowski, J. Depciuch, M. Parlińska-Wojtan, J. Baran, *Int. J. Mol. Sci.* **2018**, *19*, 4031.
- [3] Y. Li, H. Zhang, *Nanomedicine* **2019**, *14*, 1493–1512.
- [4] Y. Jabalera, F. Oltolina, M. Prat, C. Jimenez-Lopez, J. F. Fernández-Sánchez, D. Choquesillo-Lazarte, J. Gómez-Morales, *Nanomaterials* **2020**, *10*, 199.
- [5] X. Li, M. Jiang, S. Zeng, H. Liu, *Theranostics* **2019**, *9*, 3866–3878.

- [6] B. Ghaemi, O. Mashinchian, T. Mousavi, R. Karimi, S. Kharrazi, A. Amani, *ACS Appl. Mater. Interfaces* **2016**, *8*, 3123–3134.
- [7] C. Xin, X. Yao, B. Du, W. Yang, L. Wang, L. Ma, W. Weng, *Pharm. Res.* **2018**, *36*, 10.
- [8] L. Xu, Y. Xue, J. Xia, X. Qu, B. Lei, T. Yang, X. Zhang, N. Li, H. Zhao, M. Wang, M. Luo, C. Zhang, Y. Du, C. Yan, *Biomaterials* **2020**, *230*, 119670.
- [9] S. Zhao, R. Tian, B. Shao, Y. Feng, S. Yuan, L. Dong, L. Zhang, Z. Wang, H. You, *Nanoscale* **2020**, *12*, 695–702.
- [10] D. Wang, C. Zhao, G. Gao, L. Xu, G. Wang, P. Zhu, *Nanomaterials* **2019**, *9*, 1274.
- [11] J. Liu, Z. Xiong, J. Zhang, C. Peng, B. Klajnert-Maculewicz, M. Shen, X. Shi, *ACS Appl. Mater. Interfaces* **2019**, *11*, 15212–15221.
- [12] C. Chen, H. Ou, R. Liu, D. Ding, *Adv. Mater.* **2020**, *32*, 1806331.
- [13] Z. Xie, Y. Zhang, L. Liu, H. Weng, R. P. Mason, L. Tang, K. T. Nguyen, J.-T. Hsieh, J. Yang, *Adv. Mater.* **2014**, *26*, 4491–4496.
- [14] Z. Ma, H. Wan, W. Wang, X. Zhang, T. Uno, Q. Yang, J. Yue, H. Gao, Y. Zhong, Y. Tian, Q. Sun, Y. Liang, H. Dai, *Nano Res.* **2019**, *12*, 273–279.
- [15] K. O. Boakye-Yiadom, S. Kesse, Y. Opoku-Damoah, M. S. Filli, M. Aquib, M. M. B. Joelle, M. A. Farooq, R. Mavlyanova, F. Raza, R. Bavi, B. Wang, *Int. J. Pharm.* **2019**, *564*, 308–317.
- [16] S. Li, D. Amat, Z. Peng, S. Vanni, S. Raskin, G. De Angulo, A. M. Othman, R. M. Graham, R. M. Leblanc, *Nanoscale* **2016**, *8*, 16662–16669.
- [17] X. Ni, X. Zhang, X. Duan, H.-L. Zheng, X. S. Xue, D. Ding, *Nano Lett.* **2019**, *19*, 318–330.
- [18] C. Chen, X. Ni, H. W. Tian, Q. Liu, D. S. Guo, D. Ding, *Angew. Chem. Int. Ed.* **2020**, *59*, 10008.
- [19] P. Kumar, P. Tambe, K. M. Paknikar, V. Gajbhiye, *J. Controlled Release* **2018**, *287*, 35–57.
- [20] Y. Yang, K. Achazi, Y. Jia, Q. Wei, R. Haag, J. Li, *Langmuir* **2016**, *32*, 12453–12460.
- [21] M. Boffito, C. Pontremoli, S. Fiorilli, R. Laurano, G. Ciardelli, C. Vitale-Brovarone, *Pharmaceutics* **2019**, *11*, 501.
- [22] S. Nie, Y. Xing, G. J. Kim, J. W. Simons, *Annu. Rev. Biomed. Eng.* **2007**, *9*, 257–288.
- [23] J. Jeyaraman, A. Shukla, S. Sivakumar, *ACS Biomater. Sci. Eng.* **2016**, *2*, 1330–1340.
- [24] L. Dong, D. An, M. Gong, Y. Lu, H. L. Gao, Y. J. Xu, S.-H. Yu, *Small* **2013**, *9*, 3235–3241.
- [25] A. Bagheri, H. Arandiyani, C. Boyer, M. Lim, *Adv. Sci.* **2016**, *3*, 1500437-n/a.
- [26] J. Jeyaraman, A. Malecka, P. Billimoria, A. Shukla, B. Marandi, P. M. Patel, A. M. Jackson, S. Sivakumar *J. Mater. Chem. B* **2017**, *5*, 5251–5258.

- [27] C. R. Dhanya, J. Jeyaraman, P. A. Janeesh, A. Shukla, S. Sivakumar, A. Abraham, *RSC Adv.* **2016**, *6*, 55125–55134.
- [28] X. Yang, X. Han, Y. Zhu, *Colloids Surf. A* **2005**, *264*, 49–54.
- [29] K. V. Anita, S. Kumar, S. Anita, H. B. Bohidar, *Curr. Pharm. Biotechnol.* **2005**, *6*, 121–130.
- [30] M. Vubin, R. Vinayakan, F. B. Fernandez, A. John, A. Abraham, *J. Fluoresc.* **2017**, *27*, 669–677.
- [31] M. M. Joseph, S. R. Aravind, S. K. George, R. K. Pillai, S. Mini, T. T. Sreelekha, *Transl. Oncol.* **2014**, *7*, 525–536.
- [32] R. K. Desser, S. R. Himmelhoch, W. H. Evans, M. Januska, M. Mage, E. Shelton, *Arch. Biochem. Biophys.* **1972**, *148*, 452–465.
- [33] N. P. Singh, M. T. McCoy, R. R. Tice, E. L. Schneider, *Exp. Cell Res.* **1988**, *175*, 184–191.
- [34] A. Rammohan, G. Mishra, B. Mahaling, L. Tayal, A. Mukhopadhyay, S. Gambhir, A. Sharma, S. Sivakumar, *ACS Appl. Mater. Interfaces* **2016**, *8*, 350–362.
- [35] B. G. De Geest, N. N. Sanders, G. B. Sukhorukov, J. Demeester, S. C. De Smedt, *Chem. Soc. Rev.* **2007**, *36*, 636–649.
- [36] J. P. K. Tan, Q. Wang, K. C. Tam, *J. Controlled Release* **2008**, *128*, 248–254.
- [37] W. Huang, J. Zhang, H. C. Dorn, C. Zhang, *PLoS One* **2013**, *8*, e74679.
- [38] M. A. Pechenkin, H. Möhwald, D. V. Volodkin, *Soft Matter* **2012**, *8*, 8659–8665.
- [39] A. A. Antipov, G. B. Sukhorukov, S. Leporatti, I. L. Radtchenko, E. Donath, H. Möhwald, *Colloids Surf. A* **2002**, *198–200*, 535–541.
- [40] V. Sharma, A. Sundaramurthy, *Beilstein J. Nanotechnol.* **2020**, *11*, 508–532.
- [41] J. J. Richardson, J. W. Maina, H. Ejima, M. Hu, J. Guo, M. Y. Choy, S. T. Gunawan, L. Lybaert, C. E. Hagemeyer, B. G. De Geest, F. Caruso, *Adv. Sci.* **2015**, *2*, 1400007.
- [42] C. Nugraha, M. Bora, S. S. Venkatraman, *PLoS one*, Vol. *9*, **2014**, p. e92393.
- [43] A. D. Saratchandran, K. M. Sreekanth, K. V. Rao, N. Cherupally Krishnan Krishnan, *J. Cancer Ther.* **2011**, *2*, 666–674.
- [44] W. Yu, R. Liu, Y. Zhou, H. Gao, *ACS Cent. Sci.* **2020**, *6*, 100–116.
- [45] W. G. Roberts, G. E. Palade, *Cancer Res.* **1997**, *57*, 765–772.
- [46] J. K. Tee, L. X. Yip, E. S. Tan, S. Santitewagun, A. Prasath, P. C. Ke, H. K. Ho, D. T. Leong, *Chem. Soc. Rev.* **2019**, *48*, 5381–5407.
- [47] M. Nakajima, T. Irimura, G. L. Nicolson, *J. Cell. Biochem.* **1988**, *36*, 157–167.
- [48] K. J. Goodall, I. K. H. Poon, S. Phipps, M. D. Hulett, *PLoS One* **2014**, *9*, e109596.
- [49] A. Neve, F. P. Cantatore, N. Maruotti, A. Corrado, D. Ribatti, *BioMed Res. Int.* **2014**.
- [50] Y. C. Su, T. C. Cheng, Y. L. Leu, S. R. Roffler, J. Y. Wang, C. H. Chuang, C. H. Kao, K. C. Chen, H. E. Wang, T. L. Cheng, *Mol. Cancer Ther.* **2014**, *13*, 2852–2863.
- [51] G. Severini, L. Diana, R. Di Giovannandrea, C. Tirelli, *J. Cancer Res. Clin. Oncol.* **1995**, *121*, 61–63.
- [52] D. Gayathri Devi, T. R. Cibir, D. Ramaiah, A. Abraham, *J. Photochem. Photobiol. B* **2008**, *92*, 153–159.
- [53] V. Loria, I. Dato, F. Graziani, L. M. Biasucci, *Mediators Inflammation* **2008**, 135625.
- [54] A. A. Khan, M. A. Alsahli, A. H. Rahmani, *Med. Sci.* **2018**, *6*, 33.
- [55] O. Warburg, *Science* **1956**, *123*, 309–314.
- [56] M. I. Koukourakis, A. Giatromanolaki, E. Sivridis, K. C. Gatter, T. Trarbach, G. Folprecht, M. M. Shi, D. Lebwohl, T. Jalava, D. Laurent, G. Meinhardt, A. L. Harris, *Clin. Cancer Res.* **2011**, *17*, 4892–4900.
- [57] R. H. Pressley, H. G. Muntz, S. Falkenberg, L. W. Rice, *Gynecol. Oncol.* **1992**, *44*, 281–283.
- [58] S. Natesan, S. Badami, S. H. Dongre, A. Godavarthi, *J. Pharmacol. Sci.* **2007**, *103*, 12–23.
- [59] V. Jurisic, S. Radenkovic, G. Konjevic, *Advances in Cancer Biomarkers: From biochemistry to clinic for a critical revision* (Ed.: R. Scatena), Springer Netherlands, Dordrecht, **2015**, pp. 115–124.
- [60] J. J. Reichling, M. M. Kaplan, *Dig. Dis. Sci.* **1988**, *33*, 1601–1614.
- [61] M. W. Saif, D. Alexander, C. M. Wicox, *J. Appl. Res.* **2005**, *5*, 88–95.
- [62] S. Rubila, T. V. Ranganathan, K. M. Sakthivel, *Appl. Biochem. Biotechnol.* **2016**, *180*, 1482–1496.
- [63] S. Km, *Asian Pac. J. Cancer Prev.* **2013**, *14*, 3909–3919.
- [64] M. Mozaffarieh, A. Schoetzau, M. Sauter, M. Grieshaber, S. Orgül, O. Golubnitschaja, J. Flammer, *Mol Vis*, Vol. *14*, **2008**, pp. 1584–1588.
- [65] S. Wang, E. A. Konorev, S. Kotamraju, J. Joseph, S. Kalivendi, B. Kalyanaraman, *Appl. Biol. Chem.* **2004**, *279*, 25535–25543.
- [66] F. Suzuki, K. Hashimoto, H. Kikuchi, H. Nishikawa, H. Matsumoto, J. Shimada, M. Kawase, K. Sunaga, T. Tsuda, K. Satoh, H. Sakagami, *Anticancer Res.* **2005**, *25*, 887–893.
- [67] A. Kl, A. P. Kumar, A. A. Rk, *Asian Pac. J. Cancer Prev.* **2011**, *12*, 2829–2835.
- [68] S. Divakaran, C. Nai, *Int. J. Tumor Ther.* **2012**, *1*, 6–13.
- [69] M. F. Woodruff, M. O. Symes, *Br. J. Cancer* **1962**, *16*, 120–130.
- [70] F. C. Chesterman, J. J. Harvey, R. R. Dourmashkin, M. H. Salaman, *Cancer Res.* **1966**, *26*, 1759–1768.
- [71] J. J. Fang, Z. Y. Zhu, H. Dong, G. Q. Zheng, A. G. Teng, A. J. Liu, *Biomed. Environ. Sci.* **2014**, *27*, 17–26.
- [72] M. Ghoneum, N. K. Badr El-Din, S. M. Abdel Fattah, L. Tolentino, *J. Radiat. Res.* **2013**, *54*, 419–429.
- [73] A. Beheshti, J. Wage, J. T. McDonald, C. Lamont, M. Peluso, P. Hahnfeldt, L. Hlatky, *Oncotarget* **2015**, *6*, 35419–35432.
- [74] O. S. Wolfbeis, *Chem. Soc. Rev.* **2015**, *44*, 4743–4768.
- [75] L. Nobs, F. Buchegger, R. Gurny, E. Allémann, *J. Pharm. Sci.* **2004**, *93*, 1980–1992.
- [76] J. Yu, D. Javier, M. A. Yaseen, N. Nitin, R. Richards-Kortum, B. Anvari, M. S. Wong, *J. Am. Chem. Soc.* **2010**, *132*, 1929–1938.
- [77] R. I. Nicholson, J. M. Gee, M. E. Harper, *Eur. J. Cancer* **2001**, *37 Suppl. 4*, S9–15.
- [78] H. Xu, Y. Yu, D. Marciniak, A. K. Rishi, F. H. Sarkar, O. Kucuk, A. P. N. Majumdar, *Mol. Cancer Ther.* **2005**, *4*, 435–442.
- [79] M. R. Schneider, Y. Yarden, *Oncogene* **2016**, *35*, 2949–2960.

Manuscript received: July 29, 2020
Revised manuscript received: November 19, 2020
Accepted manuscript online: November 27, 2020
Version of record online: December 9, 2020

After publication of the Early View Version, the authors noticed an error made in Figure 7A and 7D during initial manuscript preparation. Accordingly, Figure 7 has been exchanged with a new version containing corrected images part (A) and (D). These errors do not affect any results or conclusions of the publication. – The Editors.

ESTIMATION OF THE SCALE PARAMETER OF A FAMILY OF DISTRIBUTIONS USING A NEWLY DERIVED MINIMAL SUFFICIENT STATISTIC

P. Yageen Thomas* and Anjana V.

Department of Statistics,
University of Kerala, Thiruvananthapuram,
Kerala, India.

*yageenthomas@gmail.com, anjana.vpn@gmail.com

ABSTRACT

A new class of statistics obtained by ordering the absolute values of the observations arising from absolutely continuous distributions which are symmetrically distributed about zero are introduced in this paper. The statistics generated by the above method are named as absolved order statistics (AOS) of the given sample. The association of the distribution of these statistics with the distribution of order statistics arising from the folded form of the parental density about zero is outlined. The vector of AOS is proved to be a minimal sufficient statistic for the class $\mathcal{F}_\theta^{(1)}$ of all absolutely continuous distributions which are symmetrically distributed about zero. A method of estimation of the scale parameter of any distribution belonging to $\mathcal{F}_\theta^{(1)}$ using AOS is described. Illustration on the advantage of the above method of estimation is described for the distributions such as (i) logistic, (ii) normal and (iii) double Weibull. A more realistic censoring scheme involving AOS as well is discussed in this paper. We have derived the U-statistic estimator based on AOS for the scale parameter σ of any distribution $f(x, \sigma) \in \mathcal{F}_\theta^{(1)}$ using the best linear unbiased estimate (BLUE) based on AOS of a preliminary sample as kernel. We have illustrated the performance of this estimator with an U-statistic generated from BLUE based on order statistics for each of (i) logistic (ii) normal and (iii) double Weibull distributions.

Key Words: Absolved Order Statistics; Symmetric Distributions; Minimal Sufficiency; Scale Parameter; Best Linear Unbiased Estimate; Censored Samples; U-statistics.

1. INTRODUCTION

Let $T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$ be the statistic based on the order statistics of a random sample of observations X_1, X_2, \dots, X_n arising from a distribution function $F \in \mathcal{F}_\theta$, where \mathcal{F}_θ is the family of all absolutely continuous univariate distributions. Then, Lehmann and Scheffe (1950) proved that T is sufficient and complete for \mathcal{F}_θ . Clearly, any statistic T induces a partition in the sample space. A sufficient statistic T is then said to be minimal sufficient if any further effort made to define a statistic T' with reduced partition of the sample space, then T' loses the property of a sufficient statistic. In particular by the use of a minimal sufficient statistic, maximum possible reduction of a data is achieved without losing any information about the parameter involved in the population distribution.

One advantage of using a sufficient statistic T in inference problems is that by restricting attention to methods based on T , no information is over looked. Then the question that immediately arises is, “what shall be the advantage of a method based on a minimal sufficient statistic when compared with the method developed for the same problem using a simply sufficient statistic ?” We discuss now the question of comparison raised above through the following example.

Suppose X_1, X_2, \dots, X_n is a random sample of size n drawn from the uniform distribution over $[-\sigma, \sigma], \sigma > 0$. If we write $X_{1:n} = \text{Min}(X_1, X_2, \dots, X_n)$ and $X_{n:n} = \text{Max}(X_1, X_2, \dots, X_n)$, then as both terminals of the range of the distribution depends on the parameter σ , we can easily verify that $X_{1:n}$ and $X_{n:n}$ are jointly sufficient for σ . If X is a random variable with uniform distribution over $[-\sigma, \sigma]$ say $U[-\sigma, \sigma]$, then it is easy to verify that $Z = \frac{X}{2\sigma} + \frac{1}{2}$ follows a uniform distribution $U[0, 1]$ defined over the interval $[0, 1]$. Therefore

$$\left(\frac{X_{1:n}}{2\sigma} + \frac{1}{2}, \frac{X_{n:n}}{2\sigma} + \frac{1}{2} \right) \stackrel{d}{=} (Y_{1:n}, Y_{n:n}), \quad (1)$$

where $Y_{1:n}$ and $Y_{n:n}$ are the smallest and largest order statistics of a random sample of size n arising from $U[0, 1]$ and $\stackrel{d}{=}$ is the usual symbol of identical in distribution used to connect two random variables.

Using the property as described in (1) and using the expression for means and variances of $Y_{1:n}$ and $Y_{n:n}$ and $Cov(Y_{1:n}, Y_{n:n})$, we obtain the following

$$E(X_{1:n}) = -\frac{n-1}{n+1}\sigma, \quad E(X_{n:n}) = \frac{n-1}{n+1}\sigma,$$

$$Var(X_{1:n}) = Var(X_{n:n}) = \frac{4n\sigma^2}{(n+1)^2(n+2)},$$

and

$$Cov(X_{1:n}, X_{n:n}) = \frac{4\sigma^2}{(n+1)^2(n+2)}.$$

Then one can obtain the best linear unbiased estimate (BLUE) of the parameter σ based on the sufficient statistic $T_1 = (X_{1:n}, X_{n:n})$ and is

$$\sigma^* = \frac{n+1}{2(n-1)}(X_{n:n} - X_{1:n}),$$

with variance given by

$$Var(\sigma^*) = \frac{2\sigma^2}{(n-1)(n+2)}. \quad (2)$$

From David and Nagaraja (2003) we notice that $U[-\sigma, \sigma]$ admits a minimal sufficient statistic $T = Max(-X_{1:n}, X_{n:n})$ for σ and the estimate $\hat{\sigma}$ of σ based on T is given by

$$\hat{\sigma} = \frac{n+1}{n}T,$$

with

$$Var(\hat{\sigma}) = \frac{\sigma^2}{n(n+2)}. \quad (3)$$

From (2) and (3) we observe that $Var(\hat{\sigma}) < Var(\sigma^*)$ and hence $\hat{\sigma}$ is relatively more efficient than σ^* . Then our conclusion is that though the sufficient statistic T_1 contains the entire information of the parameter σ , however optimal methods we apply on it to estimate σ , it shed away some amount of information when compared with the use of a minimal sufficient statistic for estimating σ .

Rao et al. (1991) and Rosaiah et al. (1991) considered some distributions which are distributed symmetrically about zero and estimated the scale parameter of those distributions using the absolute values of the order statistics arising from those distributions and

observed that their estimators are relatively more efficient than the BLUE based on the respective order statistics alone. This made the authors of this paper to suspect on the usual perception that the statistic $T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$ based on order statistics is minimal sufficient for the class $\mathcal{F}_\theta^{(1)}$ of continuous univariate distributions which are symmetrically distributed about zero in spite of the fact that $\mathcal{F}_\theta^{(1)}$ is a sub-class of the class \mathcal{F}_θ of all continuous distributions. For the scale dependent Cauchy, logistic, double Weibull and such other distributions with location parameter equal to zero and about zero they are symmetrically distributed, it is not seen in the existing literature about any investigation made to find more data reducing sufficient statistics than $T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$. This motivated the authors of this paper to identify a minimal sufficient statistic for the family $\mathcal{F}_\theta^{(1)}$ of absolutely continuous distributions which are symmetrically distributed about zero and to use that minimal sufficient statistic in modifying the already designed statistical procedures with better efficiency.

The subsequent part of this paper is organized as follows. In section 2, we introduced about AOS and proved that it constitute a minimal sufficient statistic for the family $\mathcal{F}_\theta^{(1)}$ of all continuous and symmetrically distributed univariate distributions about zero. Section 3 dealt with the estimation of the scale parameter involved in any distribution belonging to $\mathcal{F}_\theta^{(1)}$ by BLUE based on AOS. The advantage of this estimate over BLUE based on order statistics is illustrated in the case of (i) logistic (ii) normal and (iii) double Weibull distributions. In section 4, we have made a discussion on the problem of estimation of the scale parameter of distributions belonging to $\mathcal{F}_\theta^{(1)}$ by applying censoring. The linear estimates of the scale parameter of logistic, normal and double Weibull distributions due to right censoring are compared in this section with estimates based on the entire AOS. In section 5, we have described the construction of U-statistics using BLUE based on AOS as kernels for estimating the scale parameter of any distribution belonging to $\mathcal{F}_\theta^{(1)}$. In this section we have also illustrated the above method of estimation by U-statistics for logistics, normal and double Weibull distributions and made a discussion on the advantage of these U-statistics when compared with the U-statistics constructed by BLUE based on classical

order statistics as kernels. The conclusions associated with the results of this paper are described in section 6.

2. MINIMAL SUFFICIENT STATISTIC FOR THE FAMILY $\mathcal{F}_\theta^{(1)}$ OF DISTRIBUTIONS

The class $\mathcal{F}_\theta^{(1)}$ of all continuous univariate distributions with reference parameter θ which may be scalar or vector valued and are symmetrically distributed about zero has its importance in testing of hypothesis problems, for example, see (Lehmann, 1959, P.149, 206, 231, 241, 257 and 343). Distributions such as normal, logistic, double exponential, double Weibull and students t with zero mean are all members of $\mathcal{F}_\theta^{(1)}$. The well known long tailed distribution known as Cauchy distribution with median zero also belongs to $\mathcal{F}_\theta^{(1)}$. The third Hagen's hypothesis on the theory of errors (see Rao and Gupta, 1989, P.14) specifies that each component of error has an equal chance of being positive or negative. Consequently, we observe that any distribution of the error random variable satisfies the above hypothesis is symmetrically distributed about zero. One important class of distributions satisfying third Hagen's hypothesis is defined by the probability density function (pdf)

$$f(x; \beta, \sigma) = \frac{\beta}{2\sigma\Gamma(\frac{1}{\beta})} e^{-(\frac{|x|}{\sigma})^\beta}, \beta > 0, \sigma > 0, -\infty < x < \infty.$$

It is very clear to notice that $f(x; \beta, \sigma) \in \mathcal{F}_\theta^{(1)}$ for all $\beta > 0, \sigma > 0$. For similar but recently described families of distributions belonging to $\mathcal{F}_\theta^{(1)}$ see also Thomas and Priya (2015, 2016).

Hence any improvement made on a statistic which induces more reduction of the partition of the sample space when compared with that of $T = (X_{1:n}, X_{2:n}, \dots, X_{n:n})$ for the family $\mathcal{F}_\theta^{(1)}$ will contribute in devising better methods to the inference problems involved in $\mathcal{F}_\theta^{(1)}$. In order to develop a minimal sufficient statistic to the family, we define the following

Definition 2.1. *Suppose X_1, X_2, \dots, X_n is a random sample of size n drawn from a distribution with pdf $f_\theta(x)$ such that $f_\theta(x) \in \mathcal{F}_\theta^{(1)}$. If we take absolute values of the observations and order them in the increasing order of magnitude as $X_{(1:n)} \leq X_{(2:n)} \leq \dots \leq X_{(n:n)}$, then we say that $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the absolved order statistics of the given sample.*

From the given random sample of observations X_1, X_2, \dots, X_n drawn from $f_\theta(x) \in \mathcal{F}_\theta^{(1)}$, if we take the absolute values $|X_1|, |X_2|, \dots, |X_n|$ then one can easily see that $(|X_1|, |X_2|, \dots, |X_n|)$ is distributed identically as a vector of observations of a random sample of size n arising from the pdf $g_\theta(x) = 2f_\theta(x)$, $x \geq 0$, which is obtained from folding the density $f_\theta(x)$ about $x = 0$. Hence the AOS $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ defined from $(|X_1|, |X_2|, \dots, |X_n|)$ is distributed identically as the vector of order statistics of a random sample of size n drawn from the folded distribution with density $g_\theta(x) = 2f_\theta(x)$, $x \geq 0$. Now we prove the following theorem.

Theorem 2.1. *Suppose $\underline{X} = (X_1, X_2, \dots, X_n)$ is a vector of a random sample of size n drawn from a distribution with pdf $f_\theta(x) \in \mathcal{F}_\theta^{(1)}$. Let $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ be a statistic based on the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ constructed from the original sample of observations. Then T is minimal sufficient for the family $\mathcal{F}_\theta^{(1)}$.*

Proof. The joint pdf of $\underline{X} = (X_1, X_2, \dots, X_n)$ is

$$L(\underline{x}, \theta) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n f_\theta(x_{i:n}).$$

Since $f_\theta(x)$ is symmetric about zero for any given set of reals (x_1, x_2, \dots, x_n) , $\prod_{i=1}^n f_\theta(x_i)$ is a constant for all $n!$ permutations (i_1, i_2, \dots, i_n) of $(1, 2, \dots, n)$ as well as 2^n ways of assigning a coefficient $(-1)^j$, $j = 1, 2$ with each of the components x_i in (x_1, x_2, \dots, x_n) .

Thus if $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ is the statistic defined by the AOS, then the conditional pdf of $\underline{X} = (X_1, X_2, \dots, X_n)$ given $T = t = (x_{(1:n)}, x_{(2:n)}, \dots, x_{(n:n)})$ is given by

$$L(\underline{X}|T = t) = \frac{1}{2^n n!},$$

which is a constant for all $f_\theta(x) \in \mathcal{F}_\theta^{(1)}$. This proves that the statistic $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ based on AOS is sufficient for the family $\mathcal{F}_\theta^{(1)}$.

We now establish the minimal sufficient property of $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$.

Let \mathcal{D}_1 be the partition of the sample space Ω determined by the statistic $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$. Let \underline{x} be a sample point so that $\underline{x} \in \Omega$. Since T is sufficient for $\mathcal{F}_\theta^{(1)}$, using the representation given in Lindgren (1962) the partition set in \mathcal{D}_1 which contains \underline{x} is determined

by

$$E = \{y/L(y, \theta) = h(x, y)L(x, \theta)\}, \quad (4)$$

where $h(x, y)$ is independent of θ and is not equal to zero except for a set E_0 of probability measure zero. In particular, E is the set of all $y \in \Omega$ for which the ratio $\frac{L(y, \theta)}{L(x, \theta)}$ is independent of θ . Since T is sufficient for $\mathcal{F}_\theta^{(1)}$, by factorization theorem we write

$$L(x, \theta) = w_1(x)g_1(t, \theta), \quad (5)$$

where w_1 is a function of x alone and the function $g_1(t, \theta)$ depends on θ and x only through $T = t$. Now suppose that $U(X)$ is any other sufficient statistic which makes a partition \mathcal{D}_2 of the sample space Ω . Then the proof of the minimal sufficiency of T follows if we prove that each set $F \in \mathcal{D}_2$ is contained in some $E \in \mathcal{D}_1$ except possibly for a set of points with probability measure zero.

Now assume that $x, y \in F$ where $F \in \mathcal{D}_2$ so that we have $U(x) = U(y)$. But $U(x)$ is sufficient statistic and hence by factorization theorem we write

$$L(x, \theta) = w_2(x)g_2(U(x), \theta), \quad (6)$$

and

$$L(y, \theta) = w_2(y)g_2(U(y), \theta), \quad (7)$$

where $w_2(x)$ and $w_2(y)$ are not zero and independent of θ .

As $U(x) = U(y)$, we can further write

$$L(x, \theta) = w_2(x)g_2(U(y), \theta), \quad (8)$$

where w_2 is independent of θ . If $w_2(x)$ is not zero, then from (6),(7) and (8) we have

$$L(y, \theta) = \frac{w_2(y)}{w_2(x)}L(x, \theta). \quad (9)$$

Clearly $\frac{w_2(y)}{w_2(x)}$ is not zero provided $w_2(y)$ is not zero. Hence if $w_2(y) \neq 0$, then on comparing (9) with (4) we can deduce that x, y belongs to the same partition set, say $E \in \mathcal{D}_1$. This

establishes that $F \subset E$ except possibly for those points \tilde{x} such that $w_2(\tilde{x}) = 0$, however for such points $L(x, \theta) = 0$ for all θ and the collection of all such points E_0 is null in the sense that it has zero probability measure.

This establishes the theorem. □

The implications of theorem 2.1 proved in this section will have far reaching effect on the already developed statistical strategies based on the order statistics for the family $\mathcal{F}_\theta^{(1)}$ as all those strategies require suitable modification for further improvement. In the next section of this paper we describe how the modification of the estimation technique on the scale parameter of any distribution belonging to $\mathcal{F}_\theta^{(1)}$ is carried out with better efficiency.

3. MORE EFFICIENT LINEAR UNBIASED ESTIMATE OF THE SCALE PARAMETER OF SOME DISTRIBUTIONS IN $\mathcal{F}_\theta^{(1)}$

Best linear unbiased estimation of the location and scale parameters of a distribution by order statistics is extensively discussed in the existing literature. For example see Balakrishnan and Cohen (1991) and David and Nagaraja (2003). Though the modified form of the BLUE's based on order statistics of the location and scale parameters of a distribution which is symmetrically distributed about the location parameter are given by the above authors, Thomas (1990) has given a more comprehensive form of the BLUE of the scale parameter of a symmetric distribution based on quasi-ranges of the sample. Rosaiah et al. (1991) have obtained the BLUE based on absolute values of order statistics of the scale parameter σ of any distribution belonging to the family $\mathcal{F}_\theta^{(1)}$ and called their estimate as "Optimum Unbiased Absolute Estimates (OUAE)" and we denote it as $\hat{\sigma}$. These authors have further observed that the Optimum Unbiased Absolute (OUA) estimator $\hat{\sigma}$ of σ for uniform $U[-\sigma, \sigma]$ and $N(0, \sigma^2)$ distributions are more efficient than the respective BLUE σ^* based on order statistics. The above nature of OUA estimator $\hat{\sigma}$ as better estimator than the BLUE σ^* based on order statistics for double exponential and double Weibull distributions each with mean zero have been further reported by Rao et al. (1991).

As all distributions considered for comparison of OUAE $\hat{\sigma}$ with the BLUE σ^* based on order statistics by Rosaiah et al. (1991) and Rao et al. (1991) are members of $\mathcal{F}_\theta^{(1)}$, intuitively their observation on the better nature of $\hat{\sigma}$ over σ^* motivated the authors of this work and hence we have proved in section 2 that the statistic $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ based on AOS form a minimal sufficient statistic for the family $\mathcal{F}_\theta^{(1)}$. Naturally the immediate implication of the above result is that any inference procedure developed for $\mathcal{F}_\theta^{(1)}$ based on AOS will be naturally better than that based on the order statistics of the observations of a sample. The inference procedures based on absolute values of order statistics used by Rosaiah et al. (1991) and Rao et al. (1991) as well have to be generally inferior to the procedure based on the AOS of the sample, since the statistic T based on AOS as well induces a partition of Ω which is having more reduction than the partition induced by the statistic $(|X_{1:n}|, |X_{2:n}|, \dots, |X_{n:n}|)$ used by the above authors. In this section we derive the expression for the BLUE $\hat{\sigma}$ of the scale parameter σ of any distribution belonging to $\mathcal{F}_\theta^{(1)}$ based on AOS and its variance. We further illustrate how the estimate $\hat{\sigma}$ performs in a better manner than both the BLUE σ^* based on order statistics and OUAE $\hat{\sigma}$ for samples drawn from (i) the scale parameter dependent logistic distribution with zero mean (ii) normal distribution $N(0, \sigma^2)$ and (iii) scale parameter dependent double Weibull distribution with zero mean.

3.1. BLUE OF SCALE PARAMETER BASED ON AOS

It is well known that the pdf $f(x, \sigma)$ of any distribution belonging to the family $\mathcal{F}_\theta^{(1)}$ of distributions which are symmetrically distributed about zero and depends on a scale parameter σ assumes a form $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$. Now we explain below a method by which we can use the AOS to estimate σ .

Suppose $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of a random sample of size n drawn from $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$ where $f_0(y)$ is symmetrically distributed about zero. Define $Y_{(i:n)} = \frac{X_{(i:n)}}{\sigma}$ for $i = 1, 2, \dots, n$. Then $\tilde{Y} = (Y_{(1:n)}, Y_{(2:n)}, \dots, Y_{(n:n)})'$ is the vector of variables which is distributed identically as the vector of AOS of a random sample of size n drawn from $f_0(y)$. Now to estimate σ we prove the following theorem.

Theorem 3.1. *Given the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ arising from $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty$ where $f_0(\cdot)$ is symmetric about zero. Let $\tilde{Y} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})'$ be the vector of order statistics of a random sample of size n arising from the distribution with density $g(y) = 2f_0(y), y \geq 0$ which is the folded pdf of $f_0(y)$ about $y = 0$. Then for $E(\tilde{Y}) = \tilde{\gamma} = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n:n})'$ and the dispersion matrix of \tilde{Y} given by $D(\tilde{Y}) = \mathbf{G}$, the BLUE $\hat{\sigma}$ of σ based on the vector $\tilde{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})'$ of AOS is given by*

$$\hat{\sigma} = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \tilde{\gamma}' \mathbf{G}^{-1} \tilde{X}. \quad (10)$$

The variance of $\hat{\sigma}$ is given by

$$\text{Var}(\hat{\sigma}) = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \sigma^2. \quad (11)$$

Proof. If we consider the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ of the given sample then it follows that $X_{(1:n)} \leq X_{(2:n)} \leq \dots \leq X_{(n:n)}$. As described already the vector $\tilde{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})'$ is distributed as the vector of order statistics of random sample of size n drawn from the folded (about zero) form of the given distribution with pdf $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty$. Hence it is clear that $(\frac{X_{(1:n)}}{\sigma}, \frac{X_{(2:n)}}{\sigma}, \dots, \frac{X_{(n:n)}}{\sigma})'$ is distributed identically as $\tilde{Y} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})'$ where \tilde{Y} is the vector of order statistics of a random sample of size n drawn from the folded distribution with pdf $g(y) = 2f_0(y)$, $y \geq 0$. Then for $E(\tilde{Y}) = \tilde{\gamma} = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n:n})'$ and $D(\tilde{Y}) = \mathbf{G}$, where \mathbf{G} is the dispersion matrix of \tilde{Y} we have

$$E(\tilde{X}) = \tilde{\gamma} \sigma, \quad (12)$$

and

$$D(\tilde{X}) = \mathbf{G} \sigma^2. \quad (13)$$

Clearly (12) and (13) together defines a Gauss-Markov set up and hence we have the BLUE $\hat{\sigma}$ and variance of $\hat{\sigma}$ given by

$$\hat{\sigma} = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \tilde{\gamma}' \mathbf{G}^{-1} \tilde{X}, \quad (14)$$

and

$$\text{Var}(\hat{\sigma}) = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \sigma^2.$$

This establishes the theorem. □

It is straight forward to note that the BLUE $\hat{\sigma}$ as given in (10) may also be written as

$$\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)}, \quad (15)$$

where $c_{i:n}, i = 1, 2, \dots, n$ are appropriate constants determined from (14).

Now the immediate application of the theorem 3.1 may be stated in the following manner. If we have a distribution with density $f_{\theta}(x) \in \mathcal{F}_{\theta}^{(1)}$ wherein the parameters involved in θ except the scale parameter σ are known, then obtain the BLUE $\hat{\sigma}$ of σ based on the order statistics $Z_{1:n}, Z_{2:n}, \dots, Z_{n:n}$ of a random sample of size n drawn from the folded form of pdf $f_{\theta}(x)$ around $x = 0$ by $\hat{\sigma} = \sum_{i=1}^n c_{i:n} Z_{i:n}$, in which $c_{i:n}$ are determined from $\hat{\sigma} = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \tilde{\gamma}' \mathbf{G}^{-1} \tilde{Z}$ where $\tilde{\gamma} = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n:n})'$, $\tilde{\gamma}$ and \mathbf{G} are the mean vector and dispersion matrix of the vector of order statistics arising from the standard form of the distribution corresponding to the folded pdf of $f_{\theta}(x)$ around $x = 0$. The variance of $\hat{\sigma}$ is then obtained as $Var(\hat{\sigma}) = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \sigma^2$. Then without dealing any further work on the evaluation of the means, variances and covariances of AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ arising from $f_{\theta}(x), -\infty < x < \infty$, one can obtain the BLUE $\hat{\sigma}$ of σ based on the AOS as $\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)}$ with $Var(\hat{\sigma}) = (\tilde{\gamma}' \mathbf{G}^{-1} \tilde{\gamma})^{-1} \sigma^2$. It is of interest to note that $c_{i:n}$ involved in $\hat{\sigma} = \sum_{i=1}^n c_{i:n} Z_{i:n}$ as an estimate of σ of $g_{\theta}(x) = 2f_{\theta}(x), x > 0$ is same as the $c_{i:n}$ involved in $\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)}$ as estimate of σ of $f_{\theta}(x), -\infty < x < \infty$ for $i = 1, 2, \dots, n$.

3.2. BLUE BASED ON AOS OF THE SCALE PARAMETER OF LOGISTIC DISTRIBUTION

The logistic distribution is a very important distribution used for modelling a variety of statistical data sets. Balakrishnan and Cohen (1991) explained its use in the analysis of survival data, graduation of mortality statistics and as a substitute for the normal distribution. A random variable X is said to have a logistic distribution if its pdf is given by

$$f(x, \sigma) = \frac{e^{-\frac{x}{\sigma}}}{\sigma(1 + e^{-\frac{x}{\sigma}})^2}; \quad -\infty < x < \infty, \sigma > 0. \quad (16)$$

The density function given in (16) is symmetric about zero with a scale parameter σ and hence it belongs to $\mathcal{F}_{\theta}^{(1)}$. Gupta et al. (1967) dealt with the problem of estimating σ of the

Table 1: Variances of (i) BLUE σ^* based on order statistics (ii) BLUE $\hat{\sigma}$ based on AOS (iii) OUAE $\hat{\hat{\sigma}}$ and (iv) mle $\tilde{\sigma}$ of the scale parameter of logistic distribution, the relative efficiencies : $e(\hat{\hat{\sigma}}/\sigma^*)$, $e(\hat{\sigma}/\sigma^*)$ and $e(\hat{\sigma}/\tilde{\sigma})$ for $n=2(1)20$.

Sample size n	$\sigma^{-2}Var(\sigma^*)$	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\hat{\hat{\sigma}})$	$\sigma^{-2}Var(\tilde{\sigma})$	$e(\hat{\hat{\sigma}}/\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$	$e(\hat{\sigma}/\tilde{\sigma})$
2	0.6449	0.3541	0.3559	0.3497	1.8119	1.8210	0.9876
3	0.3333	0.2355	0.2368	0.2331	1.4075	1.4155	0.9898
4	0.2254	0.1763	0.1774	0.1748	1.2706	1.2786	0.9915
5	0.1704	0.1409	0.1419	0.1399	1.2011	1.2095	0.9929
6	0.1370	0.1173	0.1181	0.1165	1.1598	1.1679	0.9932
7	0.1145	0.1005	0.1012	0.0999	1.1318	1.1402	0.9940
8	0.0984	0.0879	0.0885	0.0874	1.1120	1.1202	0.9945
9	0.0863	0.0781	0.0786	0.0777	1.0973	1.1053	0.9949
10	0.0768	0.0702	0.0707	0.0699	1.0858	1.0940	0.9957
11	0.0692	0.0638	0.0643	0.06355	1.0766	1.0844	0.9961
12	0.0630	0.0585	0.0589	0.0583	1.0693	1.0768	0.9966
13	0.0578	0.0540	0.0543	0.05382	1.0629	1.0703	0.9967
14	0.0534	0.0501	0.0504	0.04995	1.0577	1.0651	0.9970
15	0.0496	0.0468	0.0476	0.04667	1.0412	1.0605	0.9972
16	0.0463	0.0438	0.0453	0.0437	1.0221	1.0563	0.9977
17	0.0434	0.0412	0.0424	0.04113	1.0245	1.0527	0.9983
18	0.0409	0.0389	0.0394	0.03884	1.0365	1.0496	0.9985
19	0.0386	0.0369	0.0371	0.03685	1.0403	1.0468	0.9986
20	0.0365	0.0350	0.0353	0.03496	1.0351	1.0413	0.9989

logistic distribution by using order statistics. The BLUE of σ based on sample quantiles was proposed by Hassanein (1969). For some other dealings associated with estimate of the scale parameter of logistic distribution one may refer to Gupta and Gnanadesikan (1966), Weke (2007), Thomas and Sreekumar (2008) and Sreekumar and Thomas (2008).

The folded form of the pdf as given in (16) about $x = 0$ is given by the pdf

$$g_{\sigma}(x) = \frac{2}{\sigma} \frac{e^{-\frac{x}{\sigma}}}{(1 + e^{-\frac{x}{\sigma}})^2}, \quad x \geq 0, \quad \sigma > 0. \quad (17)$$

The distribution defined by the above pdf is also known as half-logistic distribution. The standard form of the pdf corresponding to (17) is then given by

$$g(y) = \frac{2e^{-y}}{(1 + e^{-y})^2}, \quad y \geq 0. \quad (18)$$

It is easy to note that for the logistic distribution there exists no one dimensional sufficient statistic and hence this inability of the dimensional reduction of the sufficient statistic

make one's effort to fail in the research of finding the uniformly minimum variance unbiased estimator (UMVUE) for σ . The maximum likelihood estimator (MLE) for σ can be obtained only by solving the likelihood equation by numerical methods. The MLE is not even unbiased for small sample cases. The usual formula for variance of MLE is adoptable only for large sample sizes. Hence there is prime importance in the development of the estimation technique of σ involved in (16) for small sample cases. Since $f(x, \sigma)$ as defined in (16) belongs to $\mathcal{F}_\theta^{(1)}$, it follows that the statistic $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ based on AOS arising from (16) is minimal sufficient, we can obtain small sample estimation of σ using the AOS. Then from the theorem 3.1, we can write the estimate of σ based on AOS as

$$\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)}$$

where $c_{i:n}, i = 1, 2, \dots, n$, are determined from the equation $\hat{\sigma} = (\underline{\gamma}' \mathbf{G}^{-1} \underline{\gamma})^{-1} \underline{\gamma}' \mathbf{G}^{-1} \underline{X}$, for $\underline{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ and $\underline{\gamma}$ and \mathbf{G} denote the mean vector and dispersion matrix respectively of a vector of order statistics of a random sample of size n drawn from the distribution defined in (18). The variance of the above estimator is then given by

$$Var(\hat{\sigma}) = (\underline{\gamma}' \mathbf{G}^{-1} \underline{\gamma})^{-1} \sigma^2.$$

Balakrishnan and Puthenpura (1986) evaluated the means, variances and covariances of order statistics arising from the half-logistic distribution as defined in (18) and using those values one can determine the constants $c_{i:n}$ for $i = 1, 2, \dots, n$ and $Var(\hat{\sigma})$ for small sample sizes, with the softwares like MATHEMATICA. However, for comparison purpose of our estimator $\hat{\sigma}$ we have evaluated independently $\sigma^{-2} Var(\hat{\sigma})$ for $n = 2(1)20$ and those values are presented in Table 1. Likewise though Gupta et al. (1967) obtained the variances of the BLUE σ^* based on order statistics arising from (16) we have also evaluated those values and tabulated the values of $\sigma^{-2} Var(\sigma^*)$ for $n = 2(1)20$ in Table 1. Further following the expression for the variance of OUA estimator $\hat{\hat{\sigma}}$ as given by Rosaiah et al. (1991), we have computed parallelly the values of $\sigma^{-2} Var(\hat{\hat{\sigma}})$ for $n = 2(1)20$ and those values as well are given in Table 1. We have computed the efficiency of our estimator $\hat{\sigma}$ relative to σ^*

defined by $e(\hat{\sigma}/\sigma^*) = \frac{Var(\sigma^*)}{Var(\hat{\sigma})}$ and efficiency of OUA estimator $\hat{\sigma}$ relative to σ^* defined by $e(\hat{\sigma}/\sigma^*) = \frac{Var(\sigma^*)}{Var(\hat{\sigma})}$ for $n = 2(1)20$ and those values are also presented in Table 1.

Balakrishnan (1991) proposed a computational method to find the maximum likelihood estimates and their asymptotic variances of the parameters of logistic distribution. The estimator $\tilde{\sigma}$ which is the maximum likelihood estimate of the scale parameter σ can be evaluated by solving the equation (19) numerically.

$$\sum_{i=1}^n \frac{x_i}{\sigma^2} [1 - 2F(\frac{x_i}{\sigma})] + \frac{n}{\sigma} = 0 \quad (19)$$

Thus, we arrived the variance of $\tilde{\sigma}$ using the inverse of Information matrix and is given by

$$Var(\tilde{\sigma}) = \frac{1}{n\Delta} \sigma^2, \quad (20)$$

where $\Delta = \frac{\mu_{2,3}^{(2)}}{3} - \frac{4}{3}\mu_{1,3} - 1$, $\mu_{2,3}^{(2)}$ is the second raw moment of the order statistic $X_{2:3}$ and $\mu_{1,3}$ is the first raw moment of $X_{1:3}$. We have computed the values of $\sigma^{-2}Var(\tilde{\sigma})$ for $n = 2(1)20$ and the values obtained are also given in Table 1. The efficiency of our estimator $\hat{\sigma}$ relative to the maximum likelihood estimator $\tilde{\sigma}$ defined by $e(\hat{\sigma}/\tilde{\sigma}) = \frac{Var(\tilde{\sigma})}{Var(\hat{\sigma})}$ is also computed and is given in Table 1.

From the Table 1, we draw the following inferences: (1) The efficiency of AOS based estimator $\hat{\sigma}$ relative to σ^* is uniformly larger than unity. (2) The relative efficiency $e(\hat{\sigma}/\sigma^*)$ is uniformly larger than the relative efficiency $e(\hat{\sigma}/\sigma^*)$ indicating further that $\hat{\sigma}$ is uniformly better than the BLUE σ^* based on order statistics as well as the OUA estimator $\hat{\sigma}$. (3) The relative efficiency of our estimate $\hat{\sigma}$ to $\tilde{\sigma}$ increases as the sample size increases and approaches unity. (4) The gain in efficiency of our estimator $\hat{\sigma}$ when compared with the BLUE σ^* for estimating the scale parameter of logistic distribution ranges from 4% to 82%.

Hence for small sample cases $\hat{\sigma}$, the BLUE based on AOS is recommended as the best method for estimating the scale parameter of logistic distribution defined in (16).

3.3. BLUE BASED ON AOS OF THE SCALE PARAMETER OF NORMAL DISTRIBUTION

The normal distribution as a member of family $\mathcal{F}_\theta^{(1)}$ is defined by the pdf

$$f(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad \sigma > 0. \quad (21)$$

Let X_1, X_2, \dots, X_n be a random sample of size n drawn from (21). Clearly $f(x, \sigma)$ as defined in (21) also belongs to the exponential family of distributions, $\sum_{i=1}^n X_i^2$ forms a single sufficient statistic and hence there is not much importance in the consideration of the sufficient statistics based on the order statistics $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ and that based on the AOS $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$. However to deal with censored data arising from (21) and to deal with outliers problem in the data, the BLUE based on order statistics has to be relied upon. For the above normal distribution also, since the partition in the sample space induced by the statistic $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ based on AOS is having more reduction than the partition induced by the statistic $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ based on order statistics, we have to analyze the advantage of using BLUE $\hat{\sigma}$ based on AOS over each of the BLUE σ^* based on order statistics and the OUA estimator $\hat{\hat{\sigma}}$.

The problem of estimating σ of (21) by BLUE based on order statistics has been dealt by several authors and Sarhan and Greenberg (1962) have tabulated all means, variances and covariances of order statistics arising from the standard normal distribution and used those results to tabulate the BLUE σ^* based on order statistics for small samples and tabulated their variances. We have used their table to obtain $\sigma^{-2}Var(\sigma^*)$ for $n = 2(1)20$ and those values are tabulated in Table 2. The folded normal distribution about $x = 0$ which is also known as half-normal distribution is defined by the pdf

$$g_\sigma(x) = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0, \quad \sigma > 0. \quad (22)$$

The standard form of the distribution corresponding to (22) is then given by

$$g(y) = \sqrt{\frac{2}{\pi}} e^{-\frac{y^2}{2}}, \quad y \geq 0. \quad (23)$$

Table 2: Variances of (i) BLUE σ^* based on order statistics (ii) BLUE $\hat{\sigma}$ based on AOS (iii) OUAE $\hat{\hat{\sigma}}$ and (iv) mle $\tilde{\sigma}$ of scale parameter in $N(0, \sigma^2)$, the relative efficiencies: $e(\hat{\hat{\sigma}}/\sigma^*)$, $e(\hat{\sigma}/\sigma^*)$ and $e(\hat{\sigma}/\tilde{\sigma})$ for $n=2(1)20$.

Sample size n	$\sigma^{-2}Var(\sigma^*)$	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\hat{\hat{\sigma}})$	$\sigma^{-2}Var(\tilde{\sigma})$	$e(\hat{\hat{\sigma}}/\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$	$e(\hat{\sigma}/\tilde{\sigma})$
2	0.5708	0.2739	0.2854	0.2500	2.0001	2.0840	0.9127
3	0.2755	0.1787	0.1874	0.1667	1.4701	1.5417	0.9328
4	0.1801	0.1323	0.1391	0.1250	1.2948	1.3613	0.9448
5	0.1333	0.1049	0.1103	0.1000	1.2085	1.2707	0.9533
6	0.1057	0.0868	0.0912	0.0833	1.1590	1.2177	0.9601
7	0.0875	0.0741	0.0776	0.0714	1.1276	1.1808	0.9636
8	0.0746	0.0646	0.0675	0.0625	1.1052	1.1548	0.9675
9	0.0650	0.0572	0.0596	0.0556	1.0906	1.1364	0.9721
10	0.0576	0.0514	0.0534	0.0500	1.0787	1.1206	0.9728
11	0.0517	0.0466	0.0483	0.0455	1.0695	1.1094	0.9764
12	0.0469	0.0426	0.0441	0.0417	1.0625	1.1009	0.9789
13	0.0429	0.0393	0.0406	0.0385	1.0563	1.0916	0.9796
14	0.0395	0.0364	0.0376	0.0357	1.0505	1.0852	0.9808
15	0.0366	0.0340	0.0350	0.0334	1.0457	1.0777	0.9823
16	0.0341	0.0318	0.0327	0.0313	1.0419	1.0723	0.9843
17	0.0320	0.0299	0.0307	0.02945	1.0410	1.0702	0.9850
18	0.0301	0.0282	0.0290	0.0278	1.0369	1.0656	0.9858
19	0.0284	0.0267	0.0274	0.02634	1.0365	1.0637	0.9865
20	0.0268	0.0253	0.0260	0.0250	1.0312	1.0593	0.9881

Now from theorem 3.1 we can write the estimate of σ based on AOS as

$$\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)}$$

where for $\underline{\gamma} = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n:n})'$ and \mathbf{G} are the mean vector and dispersion matrix of a vector of order statistics of a random sample of size n drawn from the distribution defined by (23) and $c_{i:n}, i = 1, 2, \dots, n$, are determined from the equation $\hat{\sigma} = (\underline{\gamma}' \mathbf{G}^{-1} \underline{\gamma})^{-1} \underline{\gamma}' \mathbf{G}^{-1} \underline{X}$, for $\underline{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$. The variance of the above estimator is then given by

$$Var(\hat{\sigma}) = (\underline{\gamma}' \mathbf{G}^{-1} \underline{\gamma})^{-1} \sigma^2.$$

Govindarajulu and Eisenstat (1965) computed the means, variances and covariances of order statistics arising from the standard half-normal distribution as defined in (23) and used those values to determine the constants $c_{i:n}, i = 1, 2, \dots, n$, and $Var(\hat{\sigma})$ for small sample cases. However we use the software MATHEMATICA to determine all the above values ourselves for $n = 2(1)20$. The evaluated $\sigma^{-2}Var(\hat{\sigma})$ for $n = 2(1)20$ are given in Table 2. We further used the expression given for the variance of OUA estimator $\hat{\hat{\sigma}}$ as given by Rosaiah et al. (1991) and computed the values of $\sigma^{-2}Var(\hat{\hat{\sigma}})$ for $n = 2(1)20$ and those values as well are included in Table 2. We have computed the efficiencies $e(\hat{\sigma}/\sigma^*)$ of $\hat{\sigma}$ relative to σ^* and $e(\hat{\hat{\sigma}}/\sigma^*)$ of $\hat{\hat{\sigma}}$ relative to σ^* for $n = 2(1)20$ and those values are also presented in Table 2.

The maximum likelihood estimation of normal distribution is discussed widely in literature. Johnson et al. (1994) discussed the variance of the unbiased estimator of scale parameter σ when the location parameter is known. The variance of mle $\tilde{\sigma}$ of scale parameter σ can also be obtained by reversing the information matrix and is given by

$$Var(\tilde{\sigma}) = \frac{1}{2n} \sigma^2. \quad (24)$$

The evaluated variance of mle $\sigma^{-2}Var(\tilde{\sigma})$ for $n = 2(1)20$ and efficiency of our estimate $\hat{\sigma}$ relative to $\tilde{\sigma}$ for $n = 2(1)20$ are presented in Table (2). From the table, we draw the following inferences: (1) The efficiency of AOS based estimator $\hat{\sigma}$ relative to σ^* is uniformly larger than unity. (2) The relative efficiency $e(\hat{\sigma}/\sigma^*)$ is uniformly larger than the relative efficiency $e(\hat{\hat{\sigma}}/\sigma^*)$ indicating further that $\hat{\sigma}$ is uniformly better than both the estimators σ^* and $\hat{\hat{\sigma}}$.

(3) The relative efficiency of our estimate $\hat{\sigma}$ to $\tilde{\sigma}$ increases as the sample size increases and approaches unity. (4) The efficiency of our estimator $\hat{\sigma}$ when compared with BLUE σ^* for estimating σ of $N(0, \sigma^2)$ ranges from 5.5% to 108%. Hence for small sample cases BLUE $\hat{\sigma}$ of σ based on AOS help us to estimate σ of $N(0, \sigma^2)$ more efficiently when compared with its natural competitors σ^* and $\hat{\sigma}$.

3.4. BLUE BASED ON AOS OF THE SCALE PARAMETER OF DOUBLE WEIBULL DISTRIBUTION

A random variable X is said to have a double Weibull distribution if its pdf is given by

$$f(x, \alpha, \sigma) = \frac{\alpha}{2\sigma^\alpha} |x|^{\alpha-1} e^{-\left(\frac{|x|}{\sigma}\right)^\alpha}; -\infty < x < \infty, \sigma > 0, \alpha > 0. \quad (25)$$

Clearly, the above density is constructed from the well known Weibull pdf by including its reflexion on the negative axis over $(-\infty, 0]$ with it and normalizing suitably. Clearly $f(x, \alpha, \sigma) \in \mathcal{F}_\theta^{(1)}$. Let X_1, X_2, \dots, X_n be a random sample of size n drawn from (25) with the order statistics denoted by $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ and the AOS of the sample denoted by $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$. The usual perception that the statistic $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$ is minimal sufficient for the double Weibull distribution requires modification in the light of the theorem 2.1 and thus we conclude that the statistic $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ based on AOS is minimal sufficient for the distribution defined in (25). However, to obtain the BLUE σ^* for σ based on order statistics, or to obtain the BLUE $\hat{\sigma}$ based on AOS or to derive the OUA estimator $\hat{\sigma}$, we require the restriction that α is known. Thus in this section we assume through out α is known. For some given values of α , Balakrishnan and Kocherlakota (1985) derived the BLUE's of σ involved in (25) based on order statistics of a random sample of size n and its variance for $n = 2(1)10$. Rao and Narasimham (1989) extended the works of Balakrishnan and Kocherlakota (1985) to the cases for $n = 11(1)20$. The folded form of the pdf as given in (25) about $x = 0$ is given by the pdf

$$g(x, \alpha, \sigma) = \frac{\alpha}{\sigma^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\sigma}\right)^\alpha}; x \geq 0, \sigma > 0, \alpha > 0. \quad (26)$$

Table 3: Variances of (i) BLUE σ^* based on order statistics (ii) BLUE $\hat{\sigma}$ based on AOS (iii) OUAE $\hat{\hat{\sigma}}$ and (iv) UMVUE $\tilde{\sigma}$ of scale parameter in double Weibull distribution with shape parameter $\alpha = 0.5$, the relative efficiencies: $e(\hat{\hat{\sigma}}/\sigma^*)$, $e(\hat{\sigma}/\sigma^*)$ and $e(\hat{\sigma}/\tilde{\sigma})$ for $n=2(1)20$.

Sample size n	$\sigma^{-2}Var(\sigma^*)$	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\hat{\hat{\sigma}})$	$\sigma^{-2}Var(\tilde{\sigma})$	$e(\hat{\hat{\sigma}}/\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$	$e(\hat{\sigma}/\tilde{\sigma})$
2	2.9184	2.3594	2.5000	2.3233	1.1674	1.2369	0.9847
3	1.8854	1.5203	1.6019	1.5000	1.1770	1.2401	0.9867
4	1.2609	1.1151	1.1808	1.1004	1.0678	1.1308	0.9868
5	0.9786	0.8781	0.9313	0.8667	1.0508	1.1145	0.9870
6	0.7900	0.7232	0.7673	0.7143	1.0296	1.0924	0.9877
7	0.6655	0.6143	0.6513	0.6071	1.0218	1.0833	0.9883
8	0.5734	0.5336	0.5651	0.5278	1.0147	1.0746	0.9891
9	0.5041	0.4716	0.4987	0.4667	1.0108	1.0689	0.9896
10	0.4494	0.4223	0.4459	0.4182	1.0078	1.0642	0.9903
11	0.4054	0.3824	0.4031	0.3788	1.0059	1.0604	0.9906
12	0.3692	0.3493	0.3676	0.3462	1.0043	1.0571	0.9911
13	0.3389	0.3214	0.3378	0.3187	1.0032	1.0544	0.9916
14	0.3131	0.2976	0.3123	0.2953	1.0024	1.0520	0.9923
15	0.2909	0.2770	0.2904	0.2750	1.0017	1.0502	0.9928
16	0.2717	0.2592	0.2713	0.2574	1.0013	1.0480	0.9931
17	0.2548	0.2435	0.2545	0.2419	1.0009	1.0463	0.9934
18	0.2398	0.2296	0.2397	0.2281	1.0006	1.0446	0.9935
19	0.2265	0.2172	0.2265	0.2158	1.0003	1.0431	0.9936
20	0.2146	0.2060	0.2146	0.2048	1.0001	1.0420	0.9942

The standard form of the above pdf is then given by

$$g_{\alpha}(y) = \alpha y^{\alpha-1} e^{-y^{\alpha}}; y \geq 0, \alpha > 0. \quad (27)$$

Though the variance of σ^* , the BLUE of σ based on order statistics of small samples ($n \leq 20$) arising from (25) can be obtained from Balakrishnan and Kocherlakota (1985) and Rao and Narasimham (1989), we have independently evaluated the values of $\sigma^{-2}Var(\sigma^*)$ for a given value $\alpha = 0.5, 2$ and 2.5 for $n = 2(1)20$ and are given in Tables 3, 4 and 5. Following the methodology given in Rao and Gupta (1989), we have also evaluated the values of the variances of the OUA estimator $\hat{\hat{\sigma}}$ and presented the values of $\sigma^{-2}Var(\hat{\hat{\sigma}})$ for $\alpha = 0.5, 2$ and 2.5 and $n = 2(1)20$ in Tables 3, 4 and 5. If $\underline{\gamma} = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n:n})'$ and \mathbf{G} are the mean vector and dispersion matrix respectively of a vector of order statistics of a random sample of size n drawn from the pdf (27), then using the theorem 3.1 we can determine the BLUE based

Table 4: Variances of (i) BLUE σ^* based on order statistics (ii) BLUE $\hat{\sigma}$ based on AOS (iii) OUAE $\hat{\hat{\sigma}}$ and (iv) UMVUE $\tilde{\tilde{\sigma}}$ of scale parameter in double Weibull distribution with shape parameter $\alpha = 2$ the relative efficiencies: $e(\hat{\hat{\sigma}}/\sigma^*)$, $e(\hat{\sigma}/\sigma^*)$ and $e(\hat{\sigma}/\tilde{\tilde{\sigma}})$ for $n=2(1)20$.

Sample size n	$\sigma^{-2}Var(\sigma^*)$	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\hat{\hat{\sigma}})$	$\sigma^{-2}Var(\tilde{\tilde{\sigma}})$	$e(\hat{\hat{\sigma}}/\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$	$e(\hat{\sigma}/\tilde{\tilde{\sigma}})$
2	0.5234	0.1321	0.1366	0.1316	3.8316	3.9622	0.9962
3	0.2163	0.0867	0.0902	0.0864	2.3980	2.4948	0.9965
4	0.1252	0.0645	0.0671	0.06429	1.8659	1.9411	0.9967
5	0.0853	0.0514	0.0534	0.05124	1.5974	1.6595	0.9968
6	0.0641	0.0426	0.0442	0.04247	1.4502	1.5047	0.9969
7	0.0512	0.0364	0.0377	0.0363	1.4066	1.3581	0.9973
8	0.0426	0.0318	0.0328	0.03173	1.3396	1.2988	0.9977
9	0.0365	0.0282	0.0291	0.02814	1.2943	1.2543	0.9979
10	0.0319	0.0254	0.0261	0.02535	1.2559	1.2222	0.9980
11	0.0283	0.0230	0.0237	0.02296	1.2304	1.1941	0.9983
12	0.0254	0.0211	0.0216	0.02107	1.2038	1.1759	0.9986
13	0.0231	0.0195	0.0199	0.01948	1.1846	1.1608	0.9989
14	0.0212	0.0181	0.0185	0.01809	1.1713	1.1459	0.9994
15	0.0195	0.0168	0.0172	0.0168	1.1607	1.1337	1.000
16	0.0181	0.0158	0.0161	0.0158	1.1456	1.1242	1.000
17	0.0169	0.0148	0.0151	0.0148	1.1419	1.1192	1.000
18	0.0158	0.0140	0.0143	0.014	1.1286	1.1049	1.000
19	0.0149	0.0133	0.0135	0.0133	1.1203	1.1037	1.000
20	0.0140	0.0126	0.0128	0.0126	1.1111	1.0938	1.000

Table 5: Variances of (i) BLUE σ^* based on order statistics (ii) BLUE $\hat{\sigma}$ based on AOS (iii) OUAE $\hat{\hat{\sigma}}$ and and (iv) UMVUE $\tilde{\tilde{\sigma}}$ of scale parameter in double Weibull distribution with shape parameter $\alpha = 2.5$ the relative efficiencies: $e(\hat{\hat{\sigma}}/\sigma^*)$, $e(\hat{\sigma}/\sigma^*)$ and $e(\hat{\sigma}/\tilde{\tilde{\sigma}})$ for $n=2(1)20$.

Sample size n	$\sigma^{-2}Var(\sigma^*)$	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\hat{\hat{\sigma}})$	$\sigma^{-2}Var(\tilde{\tilde{\sigma}})$	$e(\hat{\hat{\sigma}}/\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$	$e(\hat{\sigma}/\tilde{\tilde{\sigma}})$
2	0.5335	0.0869	0.0916	0.08635	6.1392	5.8242	0.9936
3	0.2047	0.0567	0.0602	0.05643	3.6102	3.4003	0.9952
4	0.1087	0.0419	0.0447	0.04172	2.5943	2.4318	0.9957
5	0.0685	0.0333	0.0354	0.03321	2.0571	1.9350	0.9973
6	0.0485	0.0276	0.0292	0.02753	1.7572	1.6609	0.9975
7	0.0372	0.0235	0.0248	0.02345	1.5829	1.500	0.9978
8	0.0302	0.0205	0.0216	0.02047	1.4732	1.3981	0.9985
9	0.0254	0.0182	0.0191	0.01818	1.3956	1.3298	0.9989
10	0.0219	0.0163	0.0171	0.01629	1.3436	1.2807	0.9993
11	0.0192	0.0148	0.0155	0.01479	1.2973	1.2387	0.9994
12	0.0172	0.0136	0.0141	0.0136	1.2647	1.2199	1.000
13	0.0155	0.0125	0.0129	0.0125	1.2400	1.2016	1.000
14	0.0141	0.0116	0.0122	0.0116	1.2155	1.1557	1.000
15	0.0129	0.0108	0.0112	0.0108	1.1944	1.1518	1.000
16	0.0120	0.0101	0.0105	0.101	1.1881	1.1429	1.000
17	0.0112	0.0095	0.0098	0.0095	1.1789	1.1428	1.000
18	0.0105	0.0090	0.00926	0.00900	1.1667	1.1339	1.000
19	0.0098	0.0085	0.0088	0.008500	1.1529	1.1136	1.000
20	0.0091	0.0081	0.0083	0.008100	1.1235	1.0963	1.000

on AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ as

$$\hat{\sigma} = \sum_{i=1}^n c_{i:n} X_{(i:n)},$$

where $c_{i:n}, i = 1, 2, \dots, n$, are determined from the equation $\hat{\sigma} = (\gamma' \mathbf{G}^{-1} \gamma)^{-1} \gamma' \mathbf{G}^{-1} \tilde{X}$, for $\tilde{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})'$. The variance of the above estimator is then given by

$$Var(\hat{\sigma}) = (\gamma' \mathbf{G}^{-1} \gamma)^{-1} \sigma^2.$$

We have evaluated $\sigma^{-2} Var(\hat{\sigma})$ for $n = 2(1)20$ and those values are presented in Tables 3, 4 and 5. We have computed the efficiencies $e(\hat{\sigma}/\sigma^*)$ of $\hat{\sigma}$ relative to σ^* and $e(\hat{\hat{\sigma}}/\sigma^*)$ of OUA estimator $\hat{\hat{\sigma}}$ relative to σ^* for $n = 2(1)20$ and are listed in Tables 3, 4 and 5.

The maximum likelihood estimator $\tilde{\sigma}$ of σ can be obtained numerically for the known values of α as the solution of the equation (28),

$$\frac{\alpha \sum_{i=1}^n |x_i|^\alpha}{\sigma^{\alpha+1}} - \frac{n\alpha}{\sigma} = 0, \quad (28)$$

and also the variance of $\tilde{\sigma}$ is given by

$$Var(\tilde{\sigma}) = \frac{1}{n\alpha^2} \sigma^2. \quad (29)$$

For a known value of α , we observed that $\tilde{\sigma}$ is not an unbiased estimate of σ . Hence the efficiency of our estimate $\hat{\sigma}$, which is an unbiased one is not comparable to $\tilde{\sigma}$. Thus in this discussion we are considering the uniformly minimum variance unbiased estimator (UMVUE) $\tilde{\tilde{\sigma}}$ of the scale parameter σ .

$S = \sum_{i=1}^n |X_i|^\alpha$ is a complete sufficient statistic for Double Weibull distribution and

$T = \frac{\Gamma(n)}{\Gamma(n+\frac{1}{\alpha})} S^{\frac{1}{\alpha}}$ is an unbiased estimate of σ .

Then by Rao-Blackwell theorem $\tilde{\tilde{\sigma}} = E(T/S) = T$ is UMVUE of σ .

The variance of UMVUE $\tilde{\tilde{\sigma}}$ is given by

$$Var(\tilde{\tilde{\sigma}}) = \left[\frac{\Gamma(n)\Gamma(n+\frac{2}{\alpha})}{(\Gamma(n+\frac{1}{\alpha}))^2} - 1 \right] \sigma^2 \quad (30)$$

In Tables 3, 4 and 5, we presented the computed values of $\sigma^{-2} Var(\sigma^*)$, $\sigma^{-2} Var(\hat{\sigma})$, $\sigma^{-2} Var(\hat{\hat{\sigma}})$ and $\sigma^{-2} Var(\tilde{\tilde{\sigma}})$ of the estimators with $\alpha = 0.5$, $\alpha = 2$ and $\alpha = 2.5$ for $n = 2(1)20$.

We have computed the efficiencies $e(\hat{\sigma}/\sigma^*)$ of $\hat{\sigma}$ relative to σ^* and $e(\hat{\sigma}/\sigma^*)$ of OUA estimator $\hat{\sigma}$ relative to σ^* and $e(\hat{\sigma}/\tilde{\sigma})$ of $\hat{\sigma}$ relative to UMVUE $\tilde{\sigma}$ for $n = 2(1)20$ and are listed in Tables 3, 4 and 5.

From the above tables , we draw the following inferences: (1) The efficiency of AOS based estimator $\hat{\sigma}$ relative to σ^* is uniformly larger than unity. (2) The relative efficiency $e(\hat{\sigma}/\sigma^*)$ is uniformly larger than the relative efficiency $e(\hat{\sigma}/\sigma^*)$ indicating further that $\hat{\sigma}$ is uniformly better than the BLUE σ^* based on order statistics as well as OUA estimator $\hat{\sigma}$. (3) (3) The relative efficiency of our estimate $\hat{\sigma}$ to $\tilde{\sigma}$ increases as the sample size increases and approaches unity.(4) The gain in efficiency of our estimator $\hat{\sigma}$ when compared with the BLUE σ^* for estimating the scale parameter of double Weibull distribution for $\alpha = 0.5$ ranges from 4% to 24%.

The above method can be similarly employed for fixing the value of α by any positive real number for the estimation of σ involved in (25). Hence our observation is that for small sample cases BLUE $\hat{\sigma}$ of σ based on AOS help us to estimate σ of double Weibull distribution more efficiently when compared with its competitors σ^* and $\hat{\sigma}$.

4. LINEAR UNBIASED ESTIMATION OF SCALE PARAMETER FROM A CENSORED SAMPLE

It is interesting to note that while sampling from a symmetric distribution, if we consider the order statistics of a sample, either the largest or the smallest observation turned out to be the probable outlier. But when we consider the case of AOS, the outlier will be the largest AOS $X_{(n:n)}$. It is of interest to note that if we consider an outlier of a distribution $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty$ as the farthest lying observation from zero, then it is uniquely defined by AOS as $X_{(n:n)}$ whereas if we depend on order statistics then it is either $X_{1:n}$ or $X_{n:n}$ and thus it makes a possibility among $X_{1:n}$ and $X_{n:n}$. Hence if one is interested to have a robust estimator of the scale parameter σ using order statistics, mostly double censoring is considered. But if we use AOS to estimate robustly the parameter σ , a right censoring on the highest AOS alone is necessary. In particular if we have the sample with

observations X_1, X_2, \dots, X_n , in which we suspect that the observations with largest k absolute values are outliers, then we have to consider the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)}$ only for developing robust statistical procedure for the parent distribution. In this case k AOS are censored on the right. The estimation procedure in such a situation of the scale parameter σ of a distribution belonging to $\mathcal{F}_\theta^{(1)}$ is discussed in the following theorem.

Theorem 4.1. *Given the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)}$ arising from $f(x) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$ where $f_0(y)$ is symmetrically distributed about zero. Let $\tilde{Y} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n-k:n})'$ be the vector of order statistics of a random sample of size n arising from the distribution with density $g(y) = 2f_0(y)$, $y \geq 0$ which is the folded pdf of $f_0(y)$ about $y = 0$. Let k observations corresponding to those with largest k absolute values are censored so that the vector \tilde{X}_k of AOS available is $\tilde{X}_k = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)})'$. Let $\tilde{\gamma}_k = (\gamma_{1:n}, \gamma_{2:n}, \dots, \gamma_{n-k:n})'$ be the correspondingly partitioned part of the expectation vector $\tilde{\gamma}$ of order statistics $\tilde{Y}_k = (Y_{1:n}, Y_{2:n}, \dots, Y_{n-k:n})'$ arising from $g(y)$. Further if \mathbf{G}_k is the correspondingly partitioned matrix of \mathbf{G} such that $D(\tilde{Y}_k) = \mathbf{G}_k$, then the BLUE $\hat{\sigma}_{k,n}$ of σ based on the censored AOS is written as*

$$\hat{\sigma}_{k,n} = (\tilde{\gamma}_k' \mathbf{G}_k^{-1} \tilde{\gamma}_k)^{-1} \tilde{\gamma}_k' \mathbf{G}_k^{-1} \tilde{X}_k \quad (31)$$

The variance of $\hat{\sigma}_k$ is given by

$$\text{Var}(\hat{\sigma}_{k,n}) = (\tilde{\gamma}_k' \mathbf{G}_k^{-1} \tilde{\gamma}_k)^{-1} \sigma^2. \quad (32)$$

The proof of the above theorem is straight forward by the applications of Gauss-Markov theorem and hence is omitted.

Note 4.1. *The BLUE $\hat{\sigma}_{k,n}$ as given in (31) may be also written as*

$$\hat{\sigma}_{k,n} = \sum_{i=1}^{n-k} c_{i:n}^{(k)} X_{(i:n)} \quad (33)$$

where $c_{i:n}^{(k)}$, $i = 1, 2, \dots, n - k$ are appropriate constants.

We illustrate the above method of estimation of scale parameter in the censored case using the normal, logistic and double Weibull ($\alpha = 0.5$) distributions for a sample size $n=10$ and $k = 0, 1, 2, \dots, 8$.

Table 6: Coefficient $c_{i:10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,10} = \sum_{i=1}^{10-k} c_{i:10}^{(k)} X_{(i:10)}$, $Var(\hat{\sigma}_{k,n})$ and efficiency $e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ relative to $\hat{\sigma}$ for the scale parameter of logistic distribution for $n = 10$.

Statistic used	Coefficient $c_{i:10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,n} = \sum_{i=1}^{10-k} c_{i:10}^{(k)} X_{(i:10)}$ given along columns.								
	$\hat{\sigma}(k=0)$	$\hat{\sigma}_{1,10}(k=1)$	$\hat{\sigma}_{2,10}(k=2)$	$\hat{\sigma}_{3,10}(k=3)$	$\hat{\sigma}_{4,10}(k=4)$	$\hat{\sigma}_{5,10}(k=5)$	$\hat{\sigma}_{6,10}(k=6)$	$\hat{\sigma}_{7,10}(k=7)$	$\hat{\sigma}_{8,10}(k=8)$
$X_{(1:10)}$	0.02034	0.0223	0.025	0.02873	0.03397	0.0416	0.05342	0.0737	0.115
$X_{(2:10)}$	0.03108	0.0341	0.03819	0.04387	0.05185	0.0635	0.08147	0.1123	2.6177
$X_{(3:10)}$	0.04182	0.04584	0.05138	0.05901	0.0697	0.0853	0.1093	1.6503	
$X_{(4:10)}$	0.05223	0.05724	0.06414	0.07362	0.0869	0.1062	1.1443		
$X_{(5:10)}$	0.06195	0.06787	0.07601	0.08716	0.1027	0.8239			
$X_{(6:10)}$	0.07051	0.0772	0.08641	0.09894	0.5969				
$X_{(7:10)}$	0.07724	0.08467	0.09458	0.4241					
$X_{(8:10)}$	0.08194	0.0893	0.2859						
$X_{(9:10)}$	0.08239	0.1719							
$X_{(10:10)}$	0.07688								
$\sigma^{-2}Var(\hat{\sigma}_{k,n})$	0.0702	0.077	0.0865	0.0996	0.1182	0.1455	0.188	0.2613	0.4117
$e(\hat{\sigma}_{k,n}/\hat{\sigma})$	1.0000	0.91168	0.8116	0.7048	0.5939	0.4825	0.3734	0.7194	0.4566

In Table 6, the coefficients $c_{i:10}^{(k)}$ for $i = 1, 2, \dots, 10 - k$, $k = 0, 1, 2, \dots, 8$ of the estimates $\hat{\sigma}_{k,n}$ of the scale parameter of logistic distribution are presented. $Var(\hat{\sigma}_{k,n})$ of the estimates are also given in Table 6. When $k=0$, we write $c_{i:n}^0 = c_{i:n}$, $i = 1, 2, \dots, n$ in (33) and obtain the complete sample BLUE $\hat{\sigma}$. We have also computed the relative efficiency $e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ with respect to the complete sample BLUE $\hat{\sigma}$ defined by $e(\hat{\sigma}_k/\hat{\sigma}) = \frac{Var(\hat{\sigma})}{Var(\hat{\sigma}_{k,n})}$ and these values are also presented in Table 6.

In the case of AOS of a sample of size $n = 10$ arising from $N(0, \sigma^2)$, the estimates $\hat{\sigma}_{k,n}$ in terms of coefficients $c_{i:10}^{(k)}$ for $i = 1, 2, \dots, 10 - k$, $k = 0, 1, 2, \dots, 8$ and $Var(\hat{\sigma}_{k,n})$ are given in Table 7. When $k=0$, we write $c_{i:n}^0 = c_{i:n}$, $i = 1, 2, \dots, n$ and obtained the complete sample BLUE $\hat{\sigma}$. We have also computed the relative efficiency $e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ with respect to the complete sample BLUE $\hat{\sigma}$ defined by $e(\hat{\sigma}_{k,n}/\hat{\sigma}) = \frac{Var(\hat{\sigma})}{Var(\hat{\sigma}_{k,n})}$ and these values are listed in Table 7.

The coefficients $c_{i:10}^{(k)}$ for $i = 1, 2, \dots, 10 - k$, $k = 0, 1, 2, \dots, 8$ of the estimates $\hat{\sigma}_{k,n}$ of the scale parameter σ of double Weibull distribution with $\alpha = 0.5$ are given in Table 8. $Var(\hat{\sigma}_{k,n})$ of the estimates are also given in Table 8. When $k=0$, we write $c_{i:n}^0 = c_{i:n}$, $i = 1, 2, \dots, n$ and obtained the complete sample BLUE $\hat{\sigma}$. We have also computed the relative efficiency

Table 7: Coefficient $c_{i:10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,n} = \sum_{i=1}^{10-k} c_{i:10}^{(k)} X_{(i:10)}$, $Var(\hat{\sigma}_{k,n})$ and efficiency $e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ relative to $\hat{\sigma}$ for normal distribution for $n = 10$

Statistic used	Coefficient $c_{i:10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,n} = \sum_{i=1}^{10-k} c_{i:10}^{(k)} X_{(i:10)}$ given along columns.								
	$\hat{\sigma}(k=0)$	$\hat{\sigma}_{1,10}(k=1)$	$\hat{\sigma}_{2,10}(k=2)$	$\hat{\sigma}_{3,10}(k=3)$	$\hat{\sigma}_{4,10}(k=4)$	$\hat{\sigma}_{5,10}(k=5)$	$\hat{\sigma}_{6,10}(k=6)$	$\hat{\sigma}_{7,10}(k=7)$	$\hat{\sigma}_{8,10}(k=8)$
$X_{(1:10)}$	0.0185	0.0218	0.0257	0.0309	0.0379	0.0478	0.0631	0.0888	0.1410
$X_{(2:10)}$	0.0290	0.0339	0.0401	0.04810	0.0589	0.0745	0.0981	0.1380	4.2242
$X_{(3:10)}$	0.0401	0.0469	0.0555	0.0666	0.0816	0.1029	0.1356	2.6987	
$X_{(4:10)}$	0.0522	0.0611	0.0722	0.0866	0.1061	0.1337	1.9078		
$X_{(5:10)}$	0.0656	0.0767	0.09059	0.1086	0.13271	1.4112			
$X_{(6:10)}$	0.0806	0.0942	0.1111	0.1328	1.0612				
$X_{(7:10)}$	0.0980	0.1143	0.1345	0.7942					
$X_{(8:10)}$	0.1192	0.1386	0.5771						
$X_{(9:10)}$	0.1481	0.3901							
$X_{(10:10)}$	0.2170								
$\sigma^{-2}Var(\hat{\sigma}_{k,n})$	0.0514	0.0602	0.0714	0.0859	0.1058	0.1343	0.1779	0.2521	0.4032
$e(\hat{\sigma}_{k,n}/\hat{\sigma})$	1.000	0.8538	0.7199	0.5983	0.4858	0.3827	0.2889	0.2039	0.1275

$e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ with respect to the complete sample BLUE $\hat{\sigma}$ defined by $e(\hat{\sigma}_{k,n}/\hat{\sigma}) = \frac{Var(\hat{\sigma})}{Var(\hat{\sigma}_{k,n})}$ and these values are also presented in Table 8.

The linear unbiased estimate of the scale parameter σ of all distributions belonging to the family $\mathcal{F}_\theta^{(1)}$ which are considered in this section are obtained from a better logically supported censoring scheme in the sense that k observations having largest absolute values are censored whereas in the usual type-II double censoring there can be no fixed specification in the left and right for an equivalent censoring. For robust statistical procedures, we observe the extreme observations which are feared as outliers, eliminate them through right censoring on the AOS and devised a methodology as stated above to estimate the scale parameter based on the resulting censored data. It is clear that this approach of censoring appears to be more meaningful than other traditional procedures available for a similar situation.

When we consider the equivalent censoring using AOS with that of the censoring using order statistics, we observe that the censored observation is possibly the same when only one extreme observation is censored. Specifically that is the case when $X_{(n:n)}$ the largest AOS is censored as illustrated in our work whereas either the order statistic $X_{1:n}$ or $X_{n:n}$ is censored based on the order statistics. Since the variance of $X_{1:n}$ and $X_{n:n}$ are identically the same

Table 8: Coefficient $c_{(i:10)}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,n} = \sum_{i=1}^{10-k} c_{(i:10)}^{(k)} X_{(i:10)}$, $Var(\hat{\sigma}_{k,n})$ and efficiency $e(\hat{\sigma}_{k,n}/\hat{\sigma})$ of $\hat{\sigma}_{k,n}$ relative to $\hat{\sigma}$ for double Weibull distribution with $\alpha = 0.5$ for $n = 10$.

Statistic used	Coefficient $c_{i:10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,n} = \sum_{i=1}^{10-k} c_{(i:10)}^{(k)} X_{(i:10)}$ given along columns.								
	$\hat{\sigma}(k=0)$	$\hat{\sigma}_{1,10}(k=1)$	$\hat{\sigma}_{2,10}(k=2)$	$\hat{\sigma}_{3,10}(k=3)$	$\hat{\sigma}_{4,10}(k=4)$	$\hat{\sigma}_{5,10}(k=5)$	$\hat{\sigma}_{6,10}(k=6)$	$\hat{\sigma}_{7,10}(k=7)$	$\hat{\sigma}_{8,10}(k=8)$
$X_{(1:10)}$	0.4736	0.5277	0.5980	0.6916	0.8213	1.0114	1.3143	1.8391	3.0911
$X_{(2:10)}$	0.3203	0.3570	0.4049	0.4688	0.5578	0.6888	0.8986	1.3059	14.0208
$X_{(3:10)}$	0.2321	0.2588	0.2938	0.3407	0.4064	0.5012	0.6624	5.8025	
$X_{(4:10)}$	0.1759	0.1963	0.2232	0.2594	0.3105	0.3905	2.8271		
$X_{(5:10)}$	0.1365	0.1525	0.1738	0.2022	0.2444	1.4824			
$X_{(6:10)}$	0.1074	0.1202	0.1374	0.1619	0.7995				
$X_{(7:10)}$	0.0843	0.0946	0.1089	0.4253					
$X_{(8:10)}$	0.0649	0.0910	0.2127						
$X_{(9:10)}$	0.0473	0.0910							
$X_{(10:10)}$	0.0250								
$\sigma^{-2}Var(\hat{\sigma}_{k,n})$	0.4223	0.4697	0.5302	0.6091	0.7160	0.8681	1.1013	1.5007	2.3341
$e(\hat{\sigma}_{k,n}/\hat{\sigma})$	1.0000	0.8993	0.7966	0.6934	0.5898	0.4865	0.3835	0.2814	0.1809

for a symmetric distribution, to compare the effectiveness of censoring based on AOS with the corresponding censoring based on a single extreme order statistic, we have computed the variances of the BLUE's $\hat{\sigma}_{1,n} = \sum_{i=1}^{n-1} c_{i:n}^{(1)} X_{(i:n)}$ and $\sigma_{1,n}^* = \sum_{i=1}^{n-1} d_{i:n}^{(1)} X_{i:n}$ of the parameter σ for the cases with a single $X_{(n:n)}$ censored and a largest order statistics $X_{n:n}$ censored respectively. These variances are tabulated once again in Table 9 for each of logistic, normal and double Weibull distributions. The corresponding relative efficiencies of the AOS censored estimate with respect to order statistic estimate viz $e(\hat{\sigma}_{1,n}/\sigma_{1,n}^*) = \frac{Var(\sigma_{1,n}^*)}{Var(\hat{\sigma}_{1,n})}$ for $n=2(1)10(2)20$ and are also included in Table 9.

From Table 9, we observe that the BLUE based on a single censored AOS is uniformly more efficient than the BLUE based on the corresponding single censored order statistic for all distributions considered. This inference again illustrates how the censoring schemes based on AOS is more realistic as well as more effective in estimating the scale parameter of distributions belonging to $\mathcal{F}_\theta^{(1)}$.

Table 9: Variances of the estimates with a single censoring viz $\sigma_{1,n}^{\wedge} = \sum_{i=1}^{n-1} c_{i:n} X_{(i:n)}^{(1)}$ and $\sigma_{1,n}^* = \sum_{i=1}^{n-1} d_{i:n}^{(1)} X_{i:n}$ and relative efficiencies $e(\sigma_{1,n}^{\wedge}/\sigma_{1,n}^*) = \frac{Var(\sigma_{1,n}^*)}{Var(\sigma_{1,n}^{\wedge})}$ for $n=2(1)10(2)20$ of logistic, normal and double Weibull distributions .

Sample Size	Logistic Distribution			Normal Distribution			Double Weibull Distribution		
	$\sigma_{1,n}^{\wedge}$	$\sigma_{1,n}^*$	$e(\sigma_{1,n}^{\wedge}/\sigma_{1,n}^*)$	$\sigma_{1,n}^{\wedge}$	$\sigma_{1,n}^*$	$e(\sigma_{1,n}^{\wedge}/\sigma_{1,n}^*)$	$\sigma_{1,n}^{\wedge}$	$\sigma_{1,n}^*$	$e(\sigma_{1,n}^{\wedge}/\sigma_{1,n}^*)$
2	0.7332	2.2898	3.1230	0.6634	2.1416	3.2282	5.0000	6.8367	1.3673
3	0.356	0.6546	1.8388	0.3098	0.5447	1.7582	2.3434	3.6718	1.2767
4	0.2343	0.3446	1.4708	0.198	0.2863	1.4460	1.5090	1.7537	1.1622
5	0.1745	0.2304	1.3203	0.1444	0.1898	1.3144	1.1074	1.2512	1.1298
6	0.1392	0.1726	1.2399	0.1132	0.1406	1.2420	0.8984	0.9619	1.0707
7	0.1157	0.1379	1.1919	0.093	0.1112	1.1957	0.7380	0.7855	1.0643
8	0.0991	0.1148	1.1584	0.0787	0.0918	1.1665	0.6218	0.6616	1.0639
9	0.0867	0.0984	1.1349	0.0683	0.078	1.1420	0.5423	0.5716	1.0541
10	0.077	0.0861	1.1182	0.0602	0.0678	1.1262	0.4868	0.5027	1.0326
12	0.063	0.0689	1.0937	0.0487	0.0536	1.1006	0.3932	0.4047	1.0292
14	0.0533	0.0574	1.0769	0.0408	0.0443	1.0858	0.3328	0.3384	1.0167
16	0.0462	0.0493	1.0671	0.0352	0.0377	1.0710	0.2870	0.2905	1.0124
18	0.0408	0.0431	1.0564	0.0309	0.0329	1.0647	0.2518	0.2545	1.0106
20	0.0365	0.0382	1.0466	0.0275	0.0291	1.0582	0.2246	0.2263	1.0076

5. U-STATISTICS USING BLUE'S BASED ON AOS AS KERNELS

Hoeffding (1948), in a fundamental paper has introduced U-statistics and its applications in statistical inference. The U-statistics are strongly consistent estimators, asymptotically normal and possess the property that if U_T is the U-statistic constructed based on an unbiased estimator T of a parameter, then $Var(U_T) \leq Var(T)$ (see, Serfling, 1980, p.176). The theory of U-statistics is considered in statistical literature as one of the top twenty breakthrough type inventions made in statistics during 20th century (see Sen, 1990). For detailed descriptions about the properties and applications of U-statistics see, Randles and Wolfe (1979) and Serfling (1980).

Suppose X_1, X_2, \dots, X_m are m independent identically distributed random variables with common cumulative distribution function $F(x, \theta)$ involving a parameter θ which is estimable. Then a function $h(X_1, X_2, \dots, X_m)$ which is symmetric in X_1, X_2, \dots, X_m and

$$E(h(X_1, X_2, \dots, X_m)) = \theta$$

is called a symmetric kernel of degree m . Now for a random sample X_1, X_2, \dots, X_n (for $n > m$) drawn from $F(x, \theta)$ the statistic defined by

$$U(X_1, X_2, \dots, X_n) = \frac{1}{\binom{n}{m}} \sum_{\beta \in \mathcal{B}} h(X_{\beta_1}, X_{\beta_2}, \dots, X_{\beta_m}), \quad (34)$$

where $\mathcal{B} = \{\beta | \beta = (\beta_1, \beta_2, \dots, \beta_m), \beta_1 < \beta_2 < \dots < \beta_m\}$ is one of the $\binom{n}{m}$ combinations of m integers chosen without replacement from the set $\{1, 2, \dots, n\}$ is known as the U-statistic defined by the kernel $h(\cdot)$. Suppose that $E(h(X_1, X_2, \dots, X_m))^2 < \infty$. Let $h(X_1, X_2, \dots, X_c, X_{c+1}, \dots, X_m)$ and $h(X_1, X_2, \dots, X_c, X_{m+1}, \dots, X_{2m-c})$ be two random variables having exactly c sample observations in common, $c = 1, 2, \dots, m$. Let $\xi_c^{(m)}$ be the covariance between these two random variables. Then Hoeffding (1948) derived the variance of the U-statistic given in (34) as

$$Var[U(X_1, X_2, \dots, X_n)] = \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \xi_c^{(m)}. \quad (35)$$

Clearly $U(X_1, X_2, \dots, X_n)$ is an unbiased estimate of θ . For a detailed survey on the optimal properties of U-statistics see Serfling (1980).

5.1. U-STATISTICS AS ESTIMATOR FOR THE SCALE PARAMETER σ OF THE DISTRIBUTIONS BELONGING TO THE FAMILY $\mathcal{F}_\theta^{(1)}$

Sreekumar and Thomas (2007) have introduced the technique of constructing U-statistic estimators for the location and scale parameters of a distribution by taking BLUE based on order statistics of those parameters as kernels. Sreekumar and Thomas (2007) have also narrated a technique of obtaining the variance of those U-statistics and illustrated their method to estimate the parameters of log-gamma distribution. Sreekumar and Thomas (2008) have used their methodology further to estimate the location and scale parameters of Type I generalized logistic distribution. This U-statistic estimation to the location and scale parameters of Type I extreme value, normal and logistic distributions are dealt with in Thomas and Sreekumar (2008). For more references on the use of above mentioned method for the estimation of parameters in distributions such as skew normal, log logistic, type III generalized logistic and symmetric beta Cauchy see Thomas and Baiju (2012, 2015) and Thomas and Priya (2015, 2016).

If the distribution under interest is one belonging to $\mathcal{F}_\theta^{(1)}$ in which only the scale parameter alone is present (others if exist are known) then it is also symmetrically distributed about zero. In this case for any density $f(x, \sigma) \in \mathcal{F}_\theta^{(1)}$ we write its form as $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty, \sigma > 0$. Let X_1, X_2, \dots, X_n be a random sample of size n drawn from $f(x, \sigma)$ so that the absolute values of the observations $|X_1|, |X_2|, \dots, |X_n|$ may be considered as a random sample of size n drawn from $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma})$, $0 < z < \infty, \sigma > 0$. Consequently, we observe that the order statistics $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ constructed from $|X_1|, |X_2|, \dots, |X_n|$ are distributed as the order statistics of a random sample of size n drawn from the folded distribution with pdf $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma})$, $0 < z < \infty$. Also from theorem 2.1 it follows that the AOS $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ constitute a minimal sufficient statistic to the family of distributions $\mathcal{F}_\theta^{(1)}$.

Hence we conclude that in order to deal with the estimation of σ based on AOS arising from $f(x, \sigma) \in \mathcal{F}_\theta^{(1)}$, it is enough to deal with the estimation of σ based on order statistics arising from the folded distribution with pdf $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma})$, $0 < z < \infty$. Then to obtain

the estimate of σ based on AOS arising out of a random sample of size n from $f(x, \sigma)$, we attempt for the corresponding estimate of σ based on order statistics of a random sample of size n arising from the pdf $g(z, \sigma)$ and replace in it each order statistic by the correspondingly ordered AOS. The variance expression for the estimates remains the same for both cases.

As a result of the logical arguments made above, in order to derive a U-statistic using BLUE based on AOS as kernel for σ involved in $f(x, \sigma)$, it is enough to derive a U-statistic using BLUE based on order statistics as kernel for σ involved in $g(z, \sigma)$ and finally replace each order statistic in the estimate by the correspondingly ordered AOS. Now we describe below the method of constructing the U-statistic for σ involved in $g(z, \sigma)$ using a BLUE based on order statistics.

Let Z_1, Z_2, \dots, Z_m be an initial random sample of size m drawn from the distribution with pdf $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma}), 0 < z < \infty, \sigma > 0$. Let $\underline{Z} = (Z_{1:m}, Z_{2:m}, \dots, Z_{m:m})'$ be the vector of order statistics obtained from Z_1, Z_2, \dots, Z_m . If $\underline{V} = (V_{1:m}, V_{2:m}, \dots, V_{m:m})'$ is considered as the vector of order statistics of a random sample of size m drawn from the pdf given by $g(z, 1) = 2f_0(z), z > 0$, then the distribution of these order statistics are independent of σ . Let $E(\underline{V}) = \underline{\nu} = (\nu_{1:m}, \nu_{2:m}, \dots, \nu_{m:m})'$ be the vector of expectation of \underline{V} and let the dispersion matrix of \underline{V} be denoted by \mathbf{G} . Then from David and Nagaraja (2003), the BLUE of σ involved in $g(z, \sigma)$ based on order statistics is given by

$$\hat{\sigma} = (\underline{\nu}' \mathbf{G}^{-1} \underline{\nu})^{-1} \underline{\nu}' \mathbf{G}^{-1} \underline{Z}, \quad (36)$$

and variance of $\hat{\sigma}$ is given by

$$Var(\hat{\sigma}) = (\underline{\nu}' \mathbf{G}^{-1} \underline{\nu})^{-1} \sigma^2. \quad (37)$$

We may write $\hat{\sigma}$ also as

$$h(Z_1, Z_2, \dots, Z_m) = c_{1:m} Z_{1:m} + c_{2:m} Z_{2:m} + \dots + c_{m:m} Z_{m:m} \quad (38)$$

where $c_{1:m}, c_{2:m}, \dots, c_{m:m}$ are constants which are determined from (36). Now from Sreeku-mar and Thomas (2007) we can easily write the U-statistic for a random sample Z_1, Z_2, \dots, Z_n

($n > m$) drawn from the distribution with pdf $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma})$ based on the kernel (38) as

$$U_n^{(m)} = \frac{1}{\binom{n}{m}} \sum_{r=1}^n \left[\sum_{j=1}^m \binom{n-r}{m-j} \binom{r-1}{j-1} c_{j:m} \right] Z_{r:n}, \quad (39)$$

where we define for non-negative integers p and q , $\binom{p}{q} = 0$ for $p < q$. If we write

$$\xi_c^{(m)} = Cov[h(Z_1, Z_2, \dots, Z_c, Z_{c+1}, \dots, Z_m), h(Z_1, Z_2, \dots, Z_c, Z_{m+1}, \dots, Z_{2m-c})], \quad (40)$$

for $c = 1, 2, \dots, m$, then the $Var(U_n^{(m)})$ is given by

$$Var[U_n^{(m)}] = \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \xi_c^{(m)}. \quad (41)$$

Clearly $\xi_m^{(m)} = Var(h(Z_1, Z_2, \dots, Z_m))$ and is given by (37).

The components $\xi_c^{(m)}$ for $c = 1, 2, \dots, m-1$ involved is $Var[U_n^{(m)}]$ though look very simple, obtaining those values exactly is somewhat computationally difficult with respect to a given kernel. However Sreekumar and Thomas (2007) have given a methodology to obtain those values and the version of that methodology to our present problem is given below.

If we put $n = m + k$ in (41), then we obtain

$$Var(U_{m+k}^{(m)}) = \frac{1}{\binom{m+k}{m}} \left\{ \binom{m}{m-k} \binom{k}{k} \xi_{m-k}^{(m)} + \binom{m}{m-k+1} \binom{k}{k-1} \xi_{m-k+1}^{(m)} + \dots + \binom{m}{m} \binom{k}{0} \xi_m^{(m)} \right\}, \quad (42)$$

for $k = 1, 2, \dots, m-1$.

On putting $n = m + k$ in (39) we obtain

$$U_{m+k}^{(m)} = \frac{1}{\binom{m+k}{m}} \left\{ \left[\sum_{j=1}^m \binom{m-k-1}{m-j} \binom{0}{j-1} c_{j:m} \right] Z_{1:m+k} + \left[\sum_{j=1}^m \binom{m+k-2}{m-j} \binom{1}{j-1} c_{j:m} \right] Z_{2:m+k} + \dots \right. \\ \left. + \left[\sum_{j=1}^m \binom{0}{m-j} \binom{m+k-1}{j-1} c_{j:m} \right] Z_{m+k:m+k} \right\}. \quad (43)$$

We can write the above equation also as

$$U_{m+k}^{(m)} = \mathbf{b}_{\underset{\sim}{m+k}}' \mathbf{Z}_{\underset{\sim}{m+k}}, \quad (44)$$

where $\mathbf{Z}_{\tilde{\mathbf{m}}+\mathbf{k}} = (Z_{1:m+k}, Z_{2:m+k}, \dots, Z_{m+k:m+k})'$ and

$$\binom{m+k}{m} \mathbf{b}'_{\tilde{\mathbf{m}}+\mathbf{k}} = \left(\sum_{j=1}^m \binom{m+k-1}{m-j} \binom{0}{j-1} c_{j:m}, \sum_{j=1}^m \binom{m+k-2}{m-j} \binom{1}{j-1} c_{j:m}, \dots, \sum_{j=1}^m \binom{0}{m-j} \binom{m+k-1}{j-1} c_{j:m} \right), \quad (45)$$

where in both (43) and (45) we define $\binom{p}{q} = 0$ for $p < q$. From (44) we write

$$\text{Var}(U_{m+k}^{(m)}) = (\mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}' \mathbf{G}_{\mathbf{m}+\mathbf{k}} \mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}) \sigma^2, \quad k = 1, 2, \dots, m-1, \quad (46)$$

where $\mathbf{G}_{\mathbf{m}+\mathbf{k}}$ is the variance-covariance matrix of the vector of order statistics of a random sample of size $m+k$ drawn from the distribution with pdf $g(z, 1) = 2f_0(z)$, $z > 0$. Now from (42) and (46) we write

$$\begin{aligned} \binom{m}{m-k} \binom{k}{k} \xi_{m-k}^{(m)} + \binom{m}{m-k+1} \binom{k}{k-1} \xi_{m-k+1}^{(m)} + \dots + \binom{m}{m-1} \binom{k}{1} \xi_{m-1}^{(m)} \\ = \binom{m+k}{m} (\mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}' \mathbf{G}_{\mathbf{m}+\mathbf{k}} \mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}) \sigma^2 - \xi_m^{(m)} \end{aligned} \quad (47)$$

for $k = 1, 2, \dots, m-1$, where we define for non-negative integer p and q , $\binom{p}{q} = 0$ for $p < q$.

The above system of equations can be written by the following matrix equation:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & \binom{m}{m-1} \binom{1}{1} \\ 0 & 0 & \dots & \binom{m}{m-2} \binom{2}{2} & \binom{m}{m-1} \binom{2}{1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \binom{m}{1} \binom{m-1}{m-1} & \binom{m}{2} \binom{m-1}{m-2} & \dots & \binom{m}{m-2} \binom{m-1}{2} & \binom{m}{m-1} \binom{m-1}{1} \end{bmatrix} \begin{bmatrix} \xi_1^{(m)} \\ \xi_2^{(m)} \\ \vdots \\ \xi_{m-1}^{(m)} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{m-1} \end{bmatrix}, \quad (48)$$

where $\omega_k = \binom{m+k}{m} (\mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}' \mathbf{G}_{\mathbf{m}+\mathbf{k}} \mathbf{b}_{\tilde{\mathbf{m}}+\mathbf{k}}) \sigma^2 - \xi_m^{(m)}$, $k = 1, 2, \dots, m-1$.

If we write H to denote the coefficient matrix of the left side of (48) and ω to denote the vector in the right side of (48), then we have

$$(\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_{m-1}^{(m)})' = H^{-1} \omega.$$

The values of $\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_{m-1}^{(m)}$ can be obtained from the above equation and together with them if use $\xi_m^{(m)}$ as given in (37) and (41) we obtain the variance of the U-statistic $U_n^{(m)}$.

Thus as a result of the theory developed in this paper we write the U-statistic using the BLUE of σ based on AOS as a kernel of degree m as $U_{n,A}^{(m)}$ which we trace out from (39) and is given by

$$U_{n,A}^{(m)} = \frac{1}{\binom{n}{m}} \sum_{r=1}^n \left[\sum_{j=1}^m \binom{n-r}{m-j} \binom{r-1}{j-1} c_{j:m} \right] X_{(r:n)}, \quad (49)$$

where $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of a random sample of size n drawn from the distribution with pdf $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty, \sigma > 0$. The variance of $U_{n,A}^{(m)}$ is given by

$$Var[U_{n,A}^{(m)}] = \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \xi_c^{(m)}, \quad (50)$$

where $\xi_c^{(m)}, c = 1, 2, \dots, m-1$ are solved from (48) and $\xi_m^{(m)}$ (the variance of the kernel) is given as in (37). It is to be noted that provided $c_{j:m}$ for $j = 1, 2, \dots, m$ are determined then for any sample size n however large may be, one can obtain explicitly an unbiased, strongly consistent and asymptotically normal type of estimator $U_{n,A}^{(m)}$ for σ and is as given in (49). Similarly if the means, variances and covariances of order statistics arising from $g(z, 1) = 2f_0(z)$ for all sample sizes from m to $2m-1$ are obtained then one can determine the exact variance of the estimator $U_{n,A}^{(m)}$ for any sample size however large it may be.

From the theory of U-statistics (see, Hoeffding, 1948) we can obtain the asymptotic variance of $U_{n,A}^{(m)}$ as

$$AV(U_{n,A}^{(m)}) = \frac{m^2}{n} \xi_1^{(m)}. \quad (51)$$

Further the statistic

$$T_{n,A} = \frac{U_{n,A}^{(m)} - \sigma}{\sqrt{\frac{m^2}{n} \xi_1^{(m)}}}, \quad (52)$$

follows a standard normal distribution as $n \rightarrow \infty$ (for details see, Hoeffding, 1948). So unlike the BLUE based on order statistics, the U-statistic $U_{n,A}^{(m)}$ based on AOS as given in (49) help us to handle testing problem as well on σ for large sample size n using (52).

Now to observe if the minimal sufficient property of AOS arising from the pdf $f(x, \sigma) = \frac{1}{\sigma} f_0(\frac{x}{\sigma})$, $-\infty < x < \infty$, for $f(x, \sigma) \in \mathcal{F}_\theta^{(1)}$ has induced any improvement on the U-statistic

$U_{n,A}^{(m)}$, we obtain the variance $Var(U_{n,O}^{(m)})$ of a similar U-statistic $U_{n,O}^{(m)}$ constructed from the BLUE based on order statistics as kernel of the same degree arising from $f(x, \sigma)$. If $Var(U_{n,A}^{(m)}) < Var(U_{n,O}^{(m)})$, then as usual we claim that $U_{n,A}^{(m)}$ is a better estimate of σ than the estimate $U_{n,O}^{(m)}$. Then the efficiency of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ is denoted by $e^{(m)}$ and is given as $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$. Similarly the asymptotic relative efficiency of $U_{n,A}^{(m)}$ when compared with $U_{n,O}^{(m)}$ is denoted by $\tilde{e}^{(m)}$ and is given as $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$, where $AV(U_{n,O}^{(m)})$ can be obtained similarly and the details are available as in Sreekumar and Thomas (2007).

In the subsequent subsections we illustrate the theory discussed above to estimate the scale parameter σ involved in each of the following : (i) logistic(ii) normal and (iii) double Weibull distributions.

5.2. U-STATISTICS FOR ESTIMATING SCALE PARAMETER OF LOGISTIC DISTRIBUTION USING BLUE'S BASED ON AOS AS KERNELS

In this section we consider BLUE's based on AOS with initial sample sizes $m = 2, 3, 4$ and 5 drawn from the logistic distribution with pdf as given in (16) and take them as kernels of degrees $2, 3, 4$ and 5 respectively. Using the estimate given by (10) and its linear representation given in (15) we derive those kernels in terms of order statistics arising from half logistic distribution as defined in (17) and are given by,

$$h(Z_1, Z_2) = 0.2919Z_{1:2} + 0.3872Z_{2:2}.$$

$$h(Z_1, Z_2, Z_3) = 0.1562Z_{1:3} + 0.2347Z_{2:3} + 0.2626Z_{3:3}.$$

$$h(Z_1, Z_2, Z_3, Z_4) = 0.0979Z_{1:4} + 0.1504Z_{2:4} + 0.1912Z_{3:4} + 0.1974Z_{4:4}.$$

$$h(Z_1, Z_2, Z_3, Z_4, Z_5) = 0.06752Z_{1:5} + 0.1041Z_{2:5} + 0.1369Z_{3:5} + 0.1592Z_{4:5} + 0.1574Z_{5:5}.$$

Then using the above coefficients of the OS arising from half logistic distribution in (49) we obtain the respective U-statistics $U_{n,A}^{(2)}$, $U_{n,A}^{(3)}$, $U_{n,A}^{(4)}$ and $U_{n,A}^{(5)}$. Further we proceed as described in section 5.1, obtain the components of variances $\xi_c^{(m)}$ for $c = 1, 2, \dots, m - 1$, $m = 2, 3, 4, 5$ and those values are tabulated in Table 10.

Table 10: Variance components $\xi_c^{(m)}$ for $c = 1(1)m$ and $m = 2(1)5$ of the U-statistic $U_{n,A}^{(m)}$ for logistic distribution

m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$
2	1	0.1765	4	1	0.04386	2	1	0.0281
	2	0.3541		2	0.0879		2	0.0562
	1	0.0782		3	0.1320	5	3	0.0844
3	2	0.1567	4	4	0.1763	4	4	0.1126
	3	0.2355		5	5	0.1409		

Using those values of $\xi_c^{(m)}$ for each of $m = 2, 3, 4, 5$ we have computed $Var(U_{n,A}^{(m)})$ for $n = 10(5)20(20)100$ and for N representing large n and are given in Table 11. As described in Thomas and Sreekumar (2008), we have also calculated $Var(U_{n,O}^{(m)})$ and efficiency $e^{(m)}$ of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ for $n = 10(5)20(20)100$ and $n = N$ and these values are presented in Table 11. The asymptotic relative efficiencies $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$ for $m = 2, 3, 4, 5$ are also calculated and are included in Table 11.

From Table 11, we see that on estimating the scale parameter of logistic distribution, all efficiencies and asymptotic efficiencies of U-statistics $U_{n,A}^{(m)}$ generated from BLUE based on AOS as kernels relative to the U-statistics $U_{n,O}^{(m)}$ generated from BLUE based on classical order statistics as kernels are greater than unity. This supports the importance of the results of this paper in detecting a minimal sufficient statistic for the family of distributions belonging to $\mathcal{F}_\theta^{(1)}$ so that concentrating on it we get better statistical inferences for the family.

Table 11: Variances of U-statistics $U_{n,A}^{(m)}$, $U_{n,O}^{(m)}$ from logistic distribution, their variances and relative efficiencies $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$ and $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$ for $m = 2(1)5$.

Sample size	Variances of U-statistics										Relative efficiencies $e^{(m)}$				
	$U_{n,A}^{(2)}$	$U_{n,O}^{(2)}$	$U_{n,A}^{(3)}$	$U_{n,A}^{(3)}$	$U_{n,A}^{(4)}$	$U_{n,A}^{(4)}$	$U_{n,O}^{(4)}$	$U_{n,A}^{(5)}$	$U_{n,O}^{(5)}$	$U_{n,O}^{(5)}$	$e^{(2)}$	$e^{(3)}$	$e^{(4)}$	$e^{(5)}$	
10	0.07063	0.07745	0.07057	0.07745	0.07032	0.07719	0.07719	0.07027	0.07701	0.07701	1.0966	1.0975	1.0977	1.0959	
15	0.04711	0.05010	0.04710	0.05010	0.04686	0.04991	0.04991	0.04676	0.04981	0.04981	1.0635	1.0637	1.0651	1.0652	
20	0.03532	0.03703	0.03529	0.03703	0.03512	0.03689	0.03689	0.03504	0.03681	0.03681	1.0484	1.0493	1.0504	1.0505	
40	0.01770	0.01812	0.01769	0.01812	0.01755	0.01806	0.01806	0.01750	0.01801	0.01801	1.0237	1.0243	1.0291	1.0291	
60	0.01180	0.01200	0.01175	0.01200	0.01169	0.01196	0.01196	0.01166	0.01193	0.01193	1.0170	1.0213	1.0231	1.0232	
80	0.00883	0.00897	0.00882	0.00897	0.00878	0.00893	0.00893	0.00876	0.00891	0.00891	1.0159	1.0170	1.0171	1.0171	
100	0.00711	0.00716	0.00709	0.00716	0.00706	0.00713	0.00713	0.00703	0.00712	0.00712	1.0070	1.0099	1.0099	1.0128	
N^*	0.7059/N	0.7101/N	0.7047/N	0.7101/N	0.7018/N	0.7075/N	0.7075/N	0.7003/N	0.7060/N	0.7060/N	$\tilde{e}^{(2)} = 1.0059$	$\tilde{e}^{(3)} = 1.0077$	$\tilde{e}^{(4)} = 1.0081$	$\tilde{e}^{(m)} = 1.0081$	

*The last row values are asymptotic variances and ARE's, $\tilde{e}^{(m)}$, $m=2(1)5$

5.3. U-STATISTICS FOR ESTIMATING SCALE PARAMETER OF NORMAL DISTRIBUTION USING BLUE'S BASED ON AOS AS KERNELS

In this section, we consider BLUE's based on AOS with initial sample sizes $m = 2, 3, 4$ and 5 drawn from the normal distribution with pdf as given in (19) and take them as kernels of degrees $2, 3, 4$ and 5 respectively. Using the estimate given by (10) and its linear representation given in (15) we derive those kernels in terms of order statistics arising from half normal distribution as defined in (20) and are given by

$$h(Z_1, Z_2) = 0.3134Z_{1:2} + 0.7564Z_{2:2}$$

$$h(Z_1, Z_2, Z_3) = 0.1549Z_{1:3} + 0.2801Z_{2:3} + 0.560203Z_{3:3}$$

$$h(Z_1, Z_2, Z_3, Z_4) = 0.0939Z_{1:4} + 0.1598Z_{2:4} + 0.2496Z_{3:4} + 0.4502Z_{4:4}$$

$$h(Z_1, Z_2, Z_3, Z_4, Z_5) = 0.0636Z_{1:5} + 0.1049Z_{2:5} + 0.1557Z_{3:5} + 0.2242Z_{4:5} + 0.3786Z_{5:5}$$

Then using the above coefficients of the order statistics arising from half normal distribution in (49), we obtain the respective U-statistics $U_{n,A}^{(2)}$, $U_{n,A}^{(3)}$, $U_{n,A}^{(4)}$ and $U_{n,A}^{(5)}$. Further we proceed as described in section 5.1, obtain the components of variances $\xi_c^{(m)}$ for $c = 1, 2, \dots, m-1$ $m = 2, 3, 4, 5$ and those values are tabulated in Table 12.

Using those values of $\xi_c^{(m)}$ for each of $m = 2, 3, 4, 5$ we have computed $\text{Var}(U_{n,A}^{(m)})$ for $n = 10(5)20(20)100$ and for N representing large n and are given in Table 13.

As described in Thomas and Sreekumar (2008), we have reevaluated $\text{Var}(U_{n,O}^{(m)})$ and efficiency $e^{(m)}$ of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ for $n = 10(5)20(20)100$ and $n = N$ and these values are also presented in Table 13. The asymptotic relative efficiencies $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$ for $m = 2, 3, 4, 5$ are also calculated and included in Table 13.

From Table 13, we see that on estimating the scale parameter of normal distribution, all efficiencies and asymptotic efficiencies of U-statistics $U_{n,A}^{(m)}$ generated from BLUE based on AOS as kernels relative to the U-statistics $U_{n,O}^{(m)}$ generated from BLUE based on classical order statistics as kernels are greater than unity. This supports the importance of the results

Table 12: Variance components $\xi_c^{(m)}$ for $c = 1(1)m$ and $m = 2(1)5$ for normal distribution

m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$
2	1	0.1233	4	1	0.03121	1	0	0.02015
	2	0.2739		2	0.0648		2	0
	1	0.05567		3	0.0982	5	3	0.0620
3	2	0.1173		4	0.1323		4	0
	3	0.1787			5	0	0.1049	

of this paper in dealing with estimation of the scale parameter σ of $N(0, \sigma^2)$ by AOS more efficiently when compared with the corresponding results available based on classical order statistics.

5.4. U-STATISTICS FOR ESTIMATING SCALE PARAMETER OF DOUBLE WEIBULL DISTRIBUTION USING BLUE'S BASED ON AOS AS KERNELS

In this section, we consider BLUE's based on AOS with initial sample sizes $m = 2, 3, 4$ and 5 drawn from the double Weibull distribution with pdf as given in (22) and take them as kernels of degrees $2, 3, 4$ and 5 respectively. As discussed in the section 3.4, to obtain the BLUE of σ , we require the restriction that α is known. Hence for $\alpha = 0.5$, using the estimate given by (10) and its linear representation given in (15), we derive those kernels in terms of order statistics arising from Weibull distribution as defined in (23) and are given by

$$h(Z_1, Z_2) = 0.5781Z_{1:2} + 0.2031Z_{2:2}.$$

$$h(Z_1, Z_2, Z_3) = 0.5590Z_{1:3} + 0.2977Z_{2:3} + 0.1189Z_{3:3}.$$

$$h(Z_1, Z_2, Z_3, Z_4) = 0.5386Z_{1:4} + 0.3172Z_{2:4} + 0.1880Z_{3:4} + 0.0814Z_{4:4}.$$

$$h(Z_1, Z_2, Z_3, Z_4, Z_5) = 0.5220Z_{1:5} + 0.3229Z_{2:5} + 0.2096Z_{3:5} + 0.1328Z_{4:5} + 0.0608Z_{5:5}.$$

Table 13: Variances of U-statistics $U_{n,A}^{(m)}$, $U_{n,O}^{(m)}$ from normal distribution, their variances and relative efficiencies $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$ for different n and m together with $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$.

Sample size	Variances of U-statistics										Relative efficiencies $e^{(m)}$			
	$U_{n,A}^{(2)}$	$U_{n,O}^{(2)}$	$U_{n,A}^{(3)}$	$U_{n,O}^{(3)}$	$U_{n,A}^{(4)}$	$U_{n,O}^{(4)}$	$U_{n,A}^{(5)}$	$U_{n,O}^{(5)}$	$e^{(2)}$	$e^{(3)}$	$e^{(4)}$	$e^{(5)}$		
10	0.0500	0.0581	0.0512	0.0581	0.0515	0.0580	0.0516	0.0578	1.1620	1.1348	1.1262	1.1202		
15	0.0332	0.0371	0.0339	0.0371	0.0340	0.0370	0.0342	0.0369	1.1175	1.0944	1.0882	1.0789		
20	0.0248	0.0272	0.0253	0.0272	0.0254	0.0272	0.0255	0.0271	1.0968	1.0751	1.0709	1.0627		
40	0.0124	0.0132	0.0126	0.0132	0.0126	0.0132	0.0127	0.0131	1.0645	1.0476	1.0476	1.0315		
60	0.0082	0.0087	0.0084	0.0087	0.0084	0.0088	0.0084	0.0088	1.0610	1.0357	1.0476	1.0476		
80	0.0062	0.0065	0.0063	0.0065	0.0063	0.0065	0.0063	0.0065	1.0484	1.0317	1.0317	1.0317		
100	0.0050	0.0052	0.0050	0.0052	0.0051	0.0052	0.0051	0.0052	1.0400	1.0400	1.0196	1.0196		
N^*	0.4932/N	0.5112/N	0.4928/N	0.5029/N	0.4818/N	0.4911/N	0.4789/N	0.4835/N	$\tilde{e}^{(2)} = 1.0365$	$\tilde{e}^{(3)} = 1.0205$	$\tilde{e}^{(4)} = 1.0193$	$\tilde{e}^{(5)} = 1.0096$		

*The last row values are asymptotic variances and ARE's, $\tilde{e}^{(m)}$, $m=2(1)5$

Table 14: Variances components $\xi_c^{(m)}$ for $c = 1(1)m$ and $m = 2(1)5$ for double Weibull distribution for $\alpha = 0.5$

m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$
2	1	1.1291	4	1	0.2632	5	1	0.1663
	2	2.3594		2	0.5371		2	0.33716
	1	0.4798		3	0.8205		3	0.5128
3	2	0.9864	4	4	1.1151	5	4	0.6929
	3	1.5203					5	0.8781

Then using the above coefficients of the order statistics arising from Weibull distribution in (49) we obtain the respective U-statistics $U_{n,A}^{(2)}$, $U_{n,A}^{(3)}$, $U_{n,A}^{(4)}$ and $U_{n,A}^{(5)}$. Further we proceed as described in section 5.1, obtain the components of variances $\xi_c^{(m)}$ for $c = 1, 2, \dots, m - 1$, $m = 2, 3, 4, 5$ and $\alpha = 0.5$ and those values are tabulated in Table 14.

For $\alpha = 0.5$, using those values of $\xi_c^{(m)}$ for each of $m = 2, 3, 4, 5$ we have computed $\text{Var}(U_{n,A}^{(m)})$ for $n = 10(5)20(20)100$ and for N representing large n and are given in Table 15.

As described in Thomas and Sreekumar (2008), we have also calculated $\text{Var}(U_{n,O}^{(m)})$ and efficiency $e^{(m)}$ of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ for $n = 10(5)20(20)100$ and $n = N$ and these values are also presented in Table 15. The asymptotic relative efficiencies $\tilde{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$ for $m = 2, 3, 4, 5$ are also calculated and included in Table 15.

From Table 15, we see that on estimating the scale parameter of double Weibull distribution, all efficiencies and asymptotic efficiencies of U-statistics $U_{n,A}^{(m)}$ generated from BLUE based on AOS as kernels relative to the U-statistics $U_{n,O}^{(m)}$ generated from BLUE based on classical order statistics as kernels are greater than unity. This supports the importance of the results of this paper in detecting a minimal sufficient statistic for the family of distributions belonging to $\mathcal{F}_\theta^{(1)}$ so that concentrating on it we get better statistical inferences for the family.

Table 15: Variances of U-statistics $U_{n,A}^{(m)}, U_{n,O}^{(m)}$ from double Weibull distribution with $\alpha = 0.5$, their asymptotic variances and relative efficiencies $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$ for different n and m together with $\hat{e}^{(m)} = \frac{AV(U_{n,O}^{(m)})}{AV(U_{n,A}^{(m)})}$.

Sample Size	Variances of U-statistics										Relative efficiencies $e^{(m)}$					
	$U_{n,A}^{(2)}$	$U_{n,O}^{(2)}$	$U_{n,A}^{(3)}$	$U_{n,O}^{(3)}$	$U_{n,A}^{(4)}$	$U_{n,O}^{(4)}$	$U_{n,A}^{(5)}$	$U_{n,O}^{(5)}$	$e^{(2)}$	$e^{(3)}$	$e^{(4)}$	$e^{(5)}$	$e^{(6)}$			
10	0.4539	0.5516	0.4372	0.5516	0.4295	0.4834	0.4260	0.4710	1.2152	1.2617	1.1255	1.1056				
15	0.3021	0.3667	0.2902	0.3667	0.2844	0.3212	0.2816	0.3119	1.2138	1.2636	1.1294	1.1076				
20	0.2264	0.2747	0.2172	0.2747	0.2126	0.2406	0.2103	0.2332	1.2133	1.2647	1.1317	1.1089				
40	0.1130	0.1371	0.1083	0.1371	0.1058	0.1201	0.1045	0.1162	1.2132	1.2659	1.1352	1.1120				
60	0.0754	0.0914	0.0721	0.0914	0.0705	0.0801	0.0696	0.0774	1.2122	1.2676	1.1362	1.1120				
80	0.0566	0.0686	0.0540	0.0685	0.0529	0.0601	0.0521	0.0580	1.2120	1.2685	1.1362	1.1132				
100	0.0453	0.0548	0.0432	0.0548	0.0422	0.0480	0.0417	0.0465	1.2097	1.2685	1.1374	1.1151				
N^*	4.5164/N	5.4736/N	4.3182/N	5.4754/N	4.2112/N	4.7973/N	4.1565/N	4.6345/N	$\hat{e}^{(2)} = 1.2119$	$\hat{e}^{(3)} = 1.2680$	$\hat{e}^{(4)} = 1.1392$	$\hat{e}^{(5)} = 1.1150$				

*The last row values are asymptotic variances and ARE's, $\hat{e}^{(m)}, m = 2(1)5$

6. CONCLUSIONS

One of the conclusions of this paper is that the Absolved Order Statistics constitute a minimal sufficient statistic for the family $\mathcal{F}_\theta^{(1)}$ of all absolutely continuous distributions which are symmetrically distributed about zero. By illustrations made on the study of BLUE of the scale parameter σ based on AOS for the distributions (i) logistic (ii) normal and (iii) double Weibull, we concluded that the BLUE based on AOS is more efficient than BLUE based on classical order statistics for distributions belonging to $\mathcal{F}_\theta^{(1)}$. Again we have estimated σ of the above distributions by U-statistics constructed from BLUE based on AOS as kernels and concluded that their U-statistics are better than U-statistics constructed from BLUE based on classical order statistics as kernels. We further concluded that application of censoring with AOS is more convenient and effective than application of censoring with order statistics.

ACKNOWLEDGMENT

The anonymous reviewer is very much thanked for many of his helpful comments.

REFERENCES

- Balakrishnan, N. (1991). *Handbook of the logistic distribution*. CRC Press, New York.
- Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics & Inference: Estimation Methods*. Academic Press, New York.
- Balakrishnan, N. and Kocherlakota, S. (1985). On the double Weibull distribution: order statistics and estimation. *Sankhyā, Series B*, 47(2):161–178.
- Balakrishnan, N. and Puthenpura, S. (1986). Best linear unbiased estimators of location and scale parameters of the half logistic distribution. *Journal of Statistical Computation and Simulation*, 25(3-4):193–204.
- David, H. and Nagaraja, H. (2003). *Order Statistics*. John Wiley & Sons, New York.
- Govindarajulu, Z. and Eisenstat, S. (1965). Best estimates of location and scale parameters of a chi (1 df) distribution, using ordered observations. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers*, 12:149–164.
- Gupta, S., Qureishi, A., and Shah, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. *Technometrics*, 9(1):43–56.
- Gupta, S. S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika*, 53(3-4):565–570.
- Hassanein, K. M. (1969). Estimation of the parameters of the logistic distribution by sample quantiles. *Biometrika*, 56(3):684–687.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19:293–325.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions*, volume 1. John Wiley & Sons, New York.

- Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. John Wiley & Sons, New York.
- Lehmann, E. L. and Scheffe, H. (1950). Completeness, similar regions and unbiased estimation,. *Sankhya*, 10(4):233–268.
- Lindgren, B. W. (1962). *Statistical Theory*. MacMillan, New York.
- Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. John Wiley & Sons, New York.
- Rao, A. V. D. and Narasimham, V. L. (1989). Linear estimation in double Weibull distribution. *Sankhyā, Series B*, 51(1):24–64.
- Rao, C. R. and Gupta, S. D. (1989). *Selected papers of CR Rao*, volume 5, page 14. Taylor & Francis, New York.
- Rao, V. A., Rao, A. V. D., and Narasimham, V. L. (1991). Optimum linear unbiased estimation of the scale parameter by absolute values of order statistics in the double exponential and the double Weibull distributions. *Communications in Statistics-Simulation and Computation*, 20(4):1139–1158.
- Rosaiah, K., Kantam, R. R. L., and Narasimham, V. L. (1991). Optimum linear unbiased estimation of scale parameter by absolute values of order statistics in symmetric distributions. *Communications in Statistics-Simulation and Computation*, 20(4):1159–1172.
- Sarhan, E. A. and Greenberg, B. G. (1962). *Contributions to Order Statistics*. John Wiley & Sons, New York.
- Sen, P. K. (1990). *Breakthroughs in Statistics : Foundations and Basic Theory*, edited by Kotz S and Johnson N L, volume I. Springer, New York.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

- Sreekumar, N. and Thomas, P. Y. (2007). Estimation of the parameters of log-gamma distribution using order statistics. *Metrika*, 66(1):115–127.
- Sreekumar, N. V. and Thomas, P. Y. (2008). Estimation of the parameters of type-I generalized logistic distribution using order statistics. *Communications in Statistics-Theory and Methods*, 37(10):1506–1524.
- Thomas, P. Y. (1990). Estimating location and scale parameters of a symmetric distribution by systematic statistics. *Journal of Indian Society of Agriculture Statistics*, XLII(2):250–256.
- Thomas, P. Y. and Baiju, K. V. (2012). Estimation of the scale parameter of skew-normal distribution using U-statistics based on order statistics. *Calcutta Statistical Association Bulletin*, 64(1-2):1–20.
- Thomas, P. Y. and Baiju, K. V. (2015). Estimation of the scale parameter of log-logistic distribution using U-statistics based on order statistics. *The Indian Association for Productivity, Quality & Reliability*, 40:111–128.
- Thomas, P. Y. and Priya, R. S. (2015). On a less cumbersome method of estimation of parameters of type III generalized logistic distribution by order statistics. *Statistica*, 75(3):291–312.
- Thomas, P. Y. and Priya, R. S. (2016). Symmetric beta-Cauchy distribution and estimation of parameters using order statistics. *Calcutta Statistical Association Bulletin*, 68(1-2):111–134.
- Thomas, P. Y. and Sreekumar, N. V. (2008). Estimation of location and scale parameters of a distribution by U-statistics based on best linear functions of order statistics. *Journal of Statistical Planning and Inference*, 138(7):2190–2200.
- Weke, P. G. O. (2007). Linear estimation of scale parameter for logistic distribution based on consecutive order statistics. *Sankhyā*, 69(4):870–884.

Spatial Aeolotropy of Urban Intensity Index Based on Buffer Gradient Analysis on Urban Sprawl of Two Metropolitan Cities of Kerala

Jayalakshmy S.S¹, Dr T.K Prasad², Mereena C.S^{3,*}

¹Department of Geography, University College, Thiruvananthapuram, India

²Department of Geography, Payyanur Campus, Kannur University, Kannur, India

³Department of Geography, Govt College Kariavattom, Kerala, India

*Corresponding author: mereena_merryland@yahoo.com

Received September 04, 2021; Revised October 09, 2021; Accepted October 18, 2021

Abstract Urbanization is integrally connected to three pillars of sustainable development, economic development, social development and environment protection. Haphazard urban expansion leads to rapid sprawl, pollution and environmental degradation together with unsustainable production and consumption pattern. A comprehensive assessment and proper methodology are inevitable for urban planning. This study is to identify spatiotemporal trends of urban expansion and intensity in two Metropolitan cities: Thiruvananthapuram and Cochin city region (Kerala). To measure the magnitude and pace of urban growth and Urban proportional Index (UPI) and Urban Intensity Index (UII) were developed. GIS based buffer analysis was adopted in this study. Each buffer zone was employed as a basic spatial unit to characterize distance dependent urban growth behaviour with their UPI and UII values for a given time period. The results indicate that two distinct phases of urbanization are discernible in these. The trends in Urban behaviour of Thiruvananthapuram and Cochin are to be given grave concern and study being the administrative as well as commercial capital of Kerala state respectively.

Keywords: *spatial aeolotropy, sustainable development, sprawl, spatiotemporal, urban proportional index, urban intensity index*

Cite This Article: Jayalakshmy S.S, Dr T.K Prasad, and Mereena C.S, "Spatial Aeolotropy of Urban Intensity Index Based on Buffer Gradient Analysis on Urban Sprawl of Two Metropolitan Cities of Kerala." *Journal of City and Development*, vol. 3, no. 2 (2021): 91-100. doi: 10.12691/jcd-3-2-4.

1. Introduction

Urbanisation in a demographic sense is defined as a process of growing population concentration whereby the proportion of the total population which is classified as urban is increasing. As far as urban growth is concerned there are three components; natural increase, net in-migration and changes in city boundaries which result in incorporation of previously rural places [1,2]. Urban areas are characterized by a central feature: they concentrate population, energy and materials, industrial and commercial activities, and buildings and infrastructure. It is not any single one of these factors, but the confluence of them that defines urban area. Indeed, although there is no uniform or globally consistent definition of "urban", most countries define urban according to a criterion pertaining to some aspect of a region's population, economy, or built infrastructure [3]. The only consistent thing about cities is that they are always changing cities since their inception have always demonstrated gradual, piecemeal change through processes of accretion, addition or demolition. This type of change can be regarded as

largely cosmetic and the underlying processes of urbanisation and the overall structure of the city remained largely unchanged [4]. To understand the mechanism of urbanisation of a particular region, it is necessary to understand the growth of its satellite towns and the trend of its urbanisation and its role in the growth of regional economy and how they contribute to development of city centre. [5]

The basic premise in using remote sensing data for change detection is that the process can identify change between two or more dates that is uncharacteristic of normal variation. [6,7,8]. Remote sensing allows retrospective viewing of earth's surface, and time-series of remote sensor data can be used to develop a historical perspective of an urban attribute or process which can help examine significant human or natural processes that act over a long time period [9].

The urban population was estimated to be 2.96 billion in 2000 and 3.77 in 2010 [10]. It was estimated that nearly 50 million people are added to the world's urban population each year. The urbanization level has almost stabilized in developed countries. Africa and Asian countries are in the process of urbanization. The expansion of cities in the 20th century has been

phenomenal, and now a majority of the world's 7 billion people are living in urban areas [11].

In general Asia's urbanization closely linked with economic development. High Income countries are more urbanized with high economic growth and Populous countries witnessed regional variation in urbanization. Asia is highly populous and has experienced sustained growth for the last 3 decades, its urbanization gave birth to some of the largest and densest cities in the world, supported by massive investments in infrastructure. Within these cities the growth is not uniform it is often periphery, spreading in to neighbouring areas and or in spontaneous and unintended squatter settlements. As of 2010, Asia had 12 out of 22 mega cities in the world and the number of megacities is expected to rise considerably by 2025, to 20 in Asia.

India shares most characteristic features of urbanisation in the developing countries. It reflects a gradual increasing trend of urbanization. Urbanization in India is as pseudo urbanization were in people arrive to cities not due to urban pull but due to rural push [12]. Indian urbanization, as a whole is dominated by large cities and metropolitan towns and it is not mainly "migration lead" but a product of demographic explosion due to natural increase in nature. Traditionally, India's urbanization pattern is an illustration of highly polarised and spatially unbalanced urban system with sharp core-periphery differences. However, recent trends show that urban industrial interaction fields are spreading through linking rural areas and also small towns around the mega cities and urban corridors are emerging along the transport routes [13].

Kerala urbanisation is unique in its spatial structure and organisation. Kerala shows a distinctive pattern of rural-urban continuum throughout the state where it is difficult to demarcate urban and rural area, it presents a different picture in the matter of urbanisation from the rest of India due to dispersed nature of settlements, growth of tertiary sector and unique rural-urban continuum manifested throughout the state. In Kerala, the main reason for urban population growth is the increase in the number of urban areas and also urbanization of the peripheral areas of the existing major urban centres. Kerala has had a relatively slow but consistent growth in its urban population and the major urbanized population is concentrated on the coastal plain. The process of urbanization in Kerala is distinctly differ from other parts of India. The major differences are 1. The lesser degree of fluctuation in urban growth 2. very high spatial dispersion of towns 3. Insignificant role of Rural to Urban migration [14].

Ernakulam is the most urbanised district in the state in terms of absolute number of urban population (2234363 population as per 2011 census). After independence the political participation with people, conscious development efforts of missionaries, organizations and grassroots level planning strategies with people's participation brings Ernakulam as a miniature of Kerala model of development. Cochin city region encompasses most of urban centres in the Ernakulam district; economically active region in the state.

Thiruvananthapuram back in the time of the formation of the state was the best and foremost developed district in

Kerala. The district is developed in the base of Tourism, leisure, IT, Agriculture and Educational sector investment. Now the city region of the district is growing through more foreign and private investment. The district needed a sector wise planning approach for the holistic development.

The present study addresses urbanisation in both spatial and temporal contexts and explores urban expansion of the Cochin and Thiruvananthapuram city region as well as their interaction in the regional urbanisation. Three sets of LANDSAT images were used and GIS based buffer system was established covering these two-metropolitan regions.

2. Study Area

Urban sprawl in two fast growing urban centres in Kerala, Cochin and Thiruvananthapuram are taken as the core of present study. Former is considered to be the Commercial capital of the state, whereas the latter is administrative capital.

Cochin is one of the largest and most important commercial and industrial centres of Kerala. Its strategic importance over the centuries is underlined by the Arabian Sea. Cochin city region is situated on the south-west coast of Indian peninsula with an administrative area of 369.72 sq. km which fall within the geographical co-ordinates 9° 47' 14"N to 10° 5' 38" N and 76° 12' 22"E to 76° 25' 50"E. The Cochin Corporation, five municipalities and nine panchayats fall in the present Cochin City Region. It is located on the south-western coastal strip of India. Cochin is inseparably linked with the wetlands of Vembanad estuary. Cochin, is a region interspersed with tidal water bodies and all developmental initiatives have to be streamlined giving due respect to the geological and ecological fingerprints of the region. Large scale projects especially in construction sector, Information Technology, transport, existence of Vallarpadam container shipment contributed to the urbanisation of Cochin city.

Thiruvananthapuram is the capital city located in the south western part of Kerala having a historical legacy of administration since 1700 AD. It was the foremost developed city in Kerala having a full-fledged infrastructure setting for the administration of the kingdom. The city situated in between north latitudes 8° 17' and 8° 54' and east longitudes 76° 4' and 77° 17'. Thiruvananthapuram corporation. Three municipalities and 26 panchayaths falls under this area. The city was built on hills by the seashore comprised of three geographical divisions. The low land is a narrow stretch comprising of shorelines, rivers and deltas. The mid land region comprise low hills and valleys adjoining Ghats and the highlands forms the eastern suburbs of the city and is called as the "city of seven hills". It was the first city to have a full-fledged draining and pipeline system from the colonial period. The city covers a population of 9,57,730 persons (2011 census). The city is mainly based on tertiary sector now it is growing as one of the major IT hubs of the state.

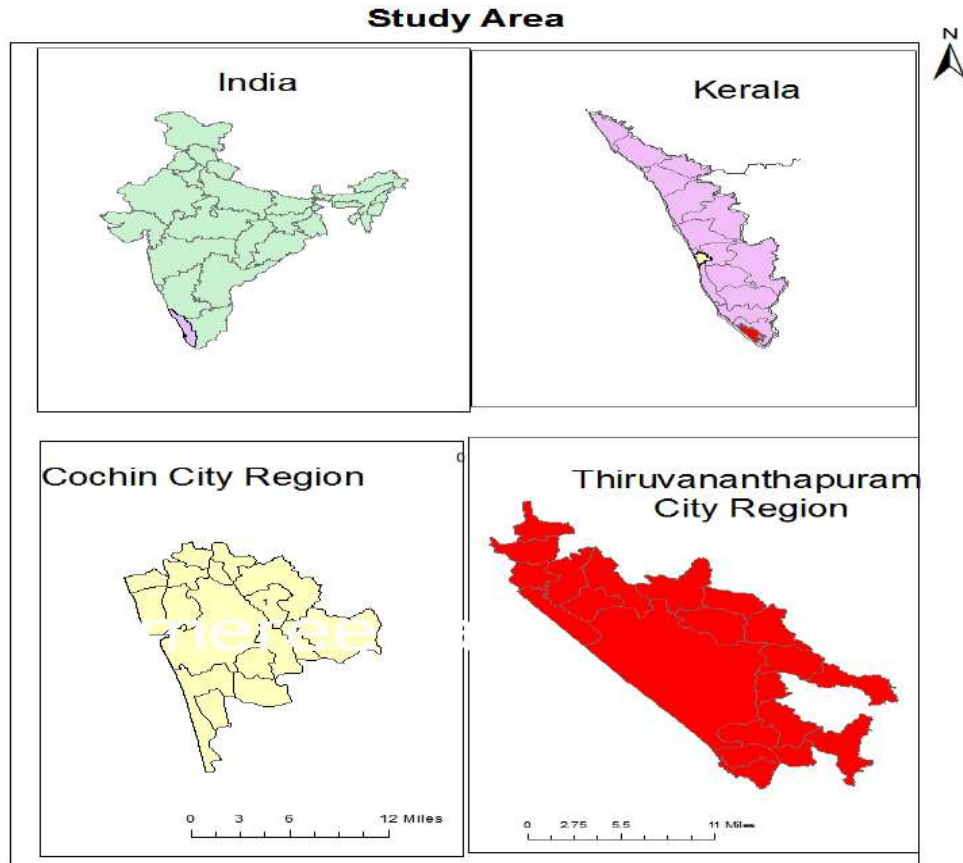


Figure 1. Location Map: Cochin & Thiruvananthapuram city region

3. Methodology

Two sets of LANDSAT TM images (1995/2005, resolution 30 m, seven bands) and one LANDSAT ETM+ (2015, resolution 15m, seven band) were used in the study. These images were processed with ERDAS IMAGINE software, which involves geometric correction, unsupervised classification and GIS reclassification. The different bands of imageries were stacked to produce a False Colour Composite (FCC). The Panchromatic data were merged with multispectral FCC. The base map (Cochin City Region Development Plan, 2031, Thiruvananthapuram Development Authority) of the study area was imported in computer environment and georeferenced in GIS environment. By using this shape files were generated. The sub-setting of satellite images was performed for extracting study area by taking geo-referenced outline Cochin City Region Development Plan boundary of 2031 map and the outline boundary of Trivandrum Development Authority. The subset image was then re-projected. All three images were then classified using ISODATA unsupervised classification algorithm. Thirty classifications. Through visual examination of satellite imageries, and by using toposheets and Google earth images, digitally classified images were interpreted and reclassified. Through the use of spectral classification, the urban area was extracted, which include high density residential areas and newly developed zone.

To measure and quantify the magnitude and pace of urban growth and Urbanisation Proportional Index (UPI) and Urbanisation Intensity Index were developed and employed [15] which expressed as

$$UPI_{i,t \sim t+n} = (ULA_{i,t+n} - ULA_{i,t}) * 100 / TLA_i$$

$$UII_{i,t \sim t+n} = \left[\frac{ULA_{i,t+n} - ULA_{i,t}}{n} \right] * 100 / TLA_i$$

The variables $UPI_{i,t \sim t+n}$, $UII_{i,t \sim t+n}$, $ULA_{i,t+n}$ and $ULA_{i,t}$ are indices of the proportion of urbanization and the intensity of urbanization within a spatial unit i during a time period $t \sim t+n$, and the areas of urban land-use for years $t+n$ and t , respectively. TLA_i is the total area of the spatial unit i . The UPI expresses the percentage of the total area occupied by urban expansion for a given spatial unit over the entire course of the study from 1995 to 2015, and it reveals the total magnitude and spatial distribution patterns of urban expansion throughout this period. The UII is used to compare the pace and intensity of urban expansion over various periods.

GIS-based buffer analysis was adopted in this study. ArcGIS software was used for buffer analysis. Each buffer zone was employed as a basic spatial unit to characterize distance-dependent urban growth behaviour with their UPI and UII values for a given time period. For the purpose of the study, three different buffer systems were established. The first was a circular buffer zone system with a buffer width of 2 km covering the entire region. This was designed to explore the overall urbanisation process over the two metropolitan regions comprising the Cochin city and Thiruvananthapuram city. The second system consists of aeolotropic buffer analysis of entire city regions. For this, the entire region was divided into eight parts based on direction. Calculations of the UPI and UII were made separately for each part to explore the directional trends in urbanisation process. The division was (in clockwise

order): north-northeast (0° to 45°), east-northeast (45° to 90°), east-southeast (90° to 135°), south-southeast (135° to 180°), South-southwest (180° to 225°), West-southwest (225° to 270°), West-northwest (270° to 315°) and North-northwest (315° to 360°). The third buffer system was established by creating separate buffer zones around Cochin and Thiruvananthapuram city centre. In all buffer analysis urban areas of Cochin and Thiruvananthapuram city region indicated by LANDSAT data from 1995, 2005 and 2015 were extracted and used to represent the urban centre as a baseline for creating buffer zones. Because the urban centre of each satellite city was generally not obvious, the CBD of each satellite city during the first-time phase (1995) was used as the origin in creating the buffer zones.

4. Analysis and Interpretation

4.1. Comparative Analysis Urban Intensity Index of Two Metropolitan Cities of Kerala: Thiruvananthapuram and Cochin

Figure 2 illustrate the changes in UII with distance from urban centre while analysing the graphs its evident that Cochin city region is already developed during 1995-2015 period but Thiruvananthapuram achieved its urban standard very recently (during 2005-2015). The Cochin city region core area well developed from the earlier days and has developed as a CBD. Well-connected transportation network added more advantage to the urbanization of this region. Between 2 to 6 km UII is increasing in this area this is because the earlier settlement of Cochin was developed at Mattanchery. This area grew in to a market town with cosmopolitan character attracting foreign traders and its commercial activities distributed along the waterfronts. In Thiruvananthapuram the city core already marked its legacy in the colonial period. The heart core of the city East fort used to be the centre of administration, art and literature from the Travancore

dynasty period. In this zone the main business centre, Chalai market is located to supply commodities to Travancore state. During 1995-2005 the well-connected road railway network accelerated the urbanization process, thus the whole sale and retailing market of the centre developed as a CBD. In Cochin city region between 4 to 8 Km UII is increasing but in Thiruvananthapuram UII is decreasing. The port-based developments in Elamkunnapuzha accelerated the urban growth. Edappaly, Vytilla and Kudannur along NH 47 developed as a active node of the city. Maradu, Thripunithura, Thrikkakara, Kakkanad, Fort cochin etc are developed in this zone and marked its high UII. In Thiruvananthapuram the UII is declining because in these zones the administrative, educational and medical servicing centres are located and the depended population is more and more commuters. Comparing with Cochin the diversified economic activities are less in Thiruvananthapuram city region it is mainly acted as a centre of administrative activity of the state while international and foreign investments are more in Cochin. The coastal belt population is also included in this region which is densely populated. A discernable UII increase recorded between 12 to 16 km the development of small marginal towns and its diversified economic activity is the main reason for this change. Balaramapuram Handloom industry, IT parks in the Kazhakootam, Kulathoor region, Development of tourist spots like Kovalam, Veli etc. added its advantages for the UII change. In Cochin city region the UII is decreasing in peripheral areas while in Thiruvananthapuram newly developing marginal towns attracts population in and surrounding areas of the city. Kattakada, Neyyatinkara and Nedumangadu are some of them developed as a marginal-towns. This region comprises of more vacant spaces with a mix of rural and urban economic activity. In Cochin the urban intensification is high in between 12km to 18 km. This zone has developed due to its geographical location and the physical development activities implemented by the government. The UII between 18-22 km shows a gradual decline as the land area is limited.

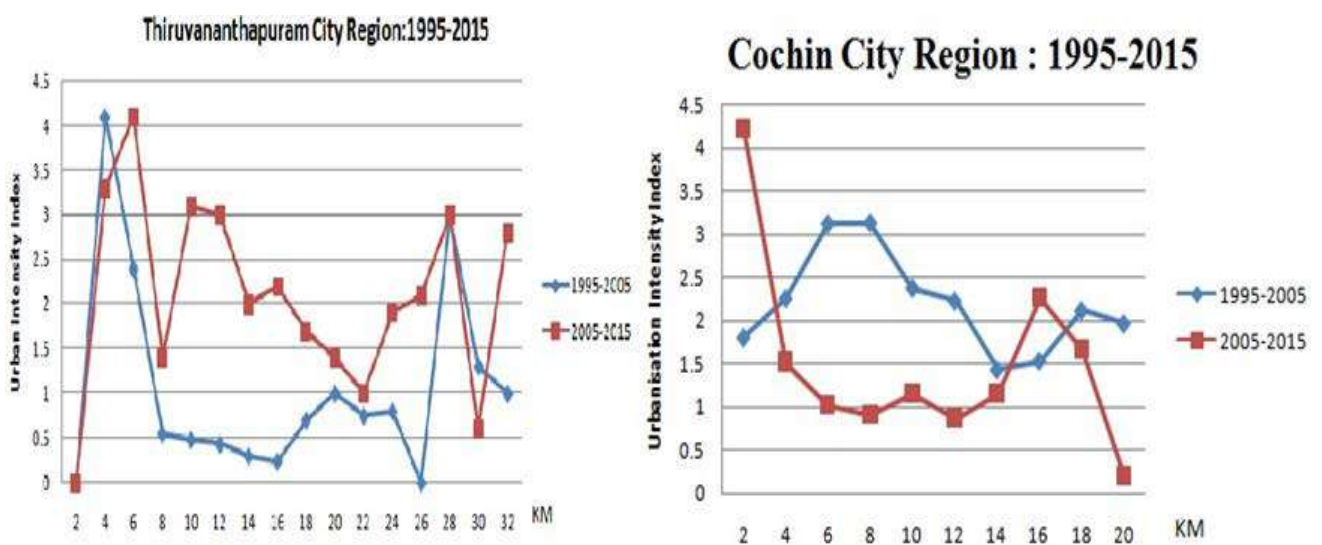


Figure 2. Changes in UII with distance from Urban centre over the Thiruvananthapuram and Cochin city region from 1995-2015

Compared to 1995-2005 period, in 2005-2015 the UII shows an increasing trend in Thiruvananthapuram but in Cochin it is decreasing. In between 2-4 km the UII is at its peak due to high intensification of Urbanization in both these cities. In between 4-8 km Cochin shows a steady declining trend from the city centre because the zone is already well developed and the vacant space for further expansion is negligible. In Thiruvananthapuram residential population increased and service-oriented activities are prominent, in between 4 to 6 km recorded sharply high UII in this zone Govt Medical College and its associated institutions located which cater a mass population not only within the district but throughout the state. The SH and MC road passing through the city enhance the developmental activities. During this period in Cochin the high UII is recorded in between 12 to 16 Km due to major investments like Brahmapuram power plant, Smart city, IT parks and related developments. In Thiruvananthapuram also a same trend of growth is evident because of IT corridor developing along the bypass Kazhakootam to Thiruvallom which is in between 10 to 14 Km. Afterwards a gradual declining trend is visible up to 14Km. In these areas some panchayaths are also included having both urban and rural characteristics and have a potential to develop as a satellite town such as Nemom, Sreekaryam, Peroorkada etc. Cochin also shows a same trend of decline in UII is recorded while the zone is mixed-up with rural economic activities.

4.2. Spatial Aeolotropy of UII Based on Buffer Analysis during 1995-2015

Figure 3, Figure 4 and Figure 5 displays the results of the spatial aeolotropy analysis of the UII for Cochin and Thiruvananthapuram city region. During 1995-2005 Cochin city up to 2 Km NNE, SSE, ESE slices shows a similar trend of high urban growth as the city core comes under this slice. ENE, SSW, shows a steady growth in UII.

WSW, WNW and NNW slices shows a gradual increase in urbanization compared to other slice as the portion largely covered by water bodies. The ENE slices which cover Ernakulum town and Edappally reaches its peak value in this buffer zone as the residential population is more concentrated in this zone. In Thiruvananthapuram NNE, WSW and WNW slices shows a similar trend of high urban growth rather than economic activities, administrative activities are predominant. Most of department headquarters of the state located in this region. ESE, SSW, WNW&NNW shows a sharp decline in between 4 to 6 KM. WSW &SSW UII decreasing and further expansion hindered by seacoast.

During 1995-2005 Thiruvananthapuram recorded comparatively low rate of urban growth. Moving apart from the city centre the UII is decreasing and in between rural and urban functions mutually coexisting. ESE and SSE slices up to 12Km shows a decreasing trend, agricultural fields are seen in this direction but these areas recorded high UII up to 16 Km. Balaramapuram (Handloom small scale industry) Neyyatinkara and Kattakada are located in this zone. These small marginal towns acted as a collection centres for the agriculture and related Products from the nearby Panchayath. In Cochin NNE, ESE and SSE demonstrates a narrow increase in UII between 4-6Km. WSW slice including Mattanchery, Bolgatty displays a high peak of UII. Similar to Thiruvananthapuram due to physiographic constraints further urban expansion is hindered in WSW direction. In between 6 to 8 km NNE, ENE and NNW slices show a dramatic reduction in UII due to the predominance of primary activities. A rapid tapering off to relatively low levels in the range 8-12 km is seen in NNE, ENE, SSE and NNW slices and further declining trend is visible on ESE and SSE direction. Contrary to this WNW slices including Vallarpadom, Njarakkal and Elamkunnappuzha UII shows an increasing trend up to 10Km. After 12 km ENE, SSE, ESE and SSW slices shows uniform growth.

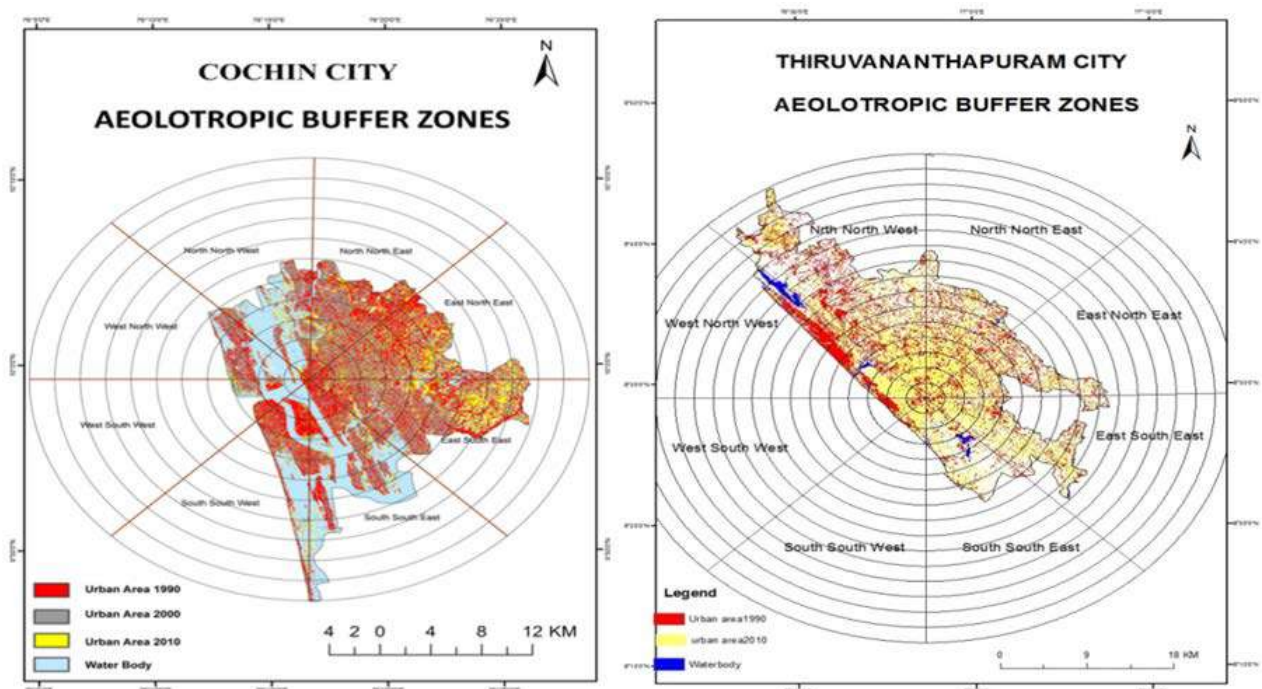


Figure 3. Aeolotropic buffer zones: Cochin and Thiruvananthapuram city

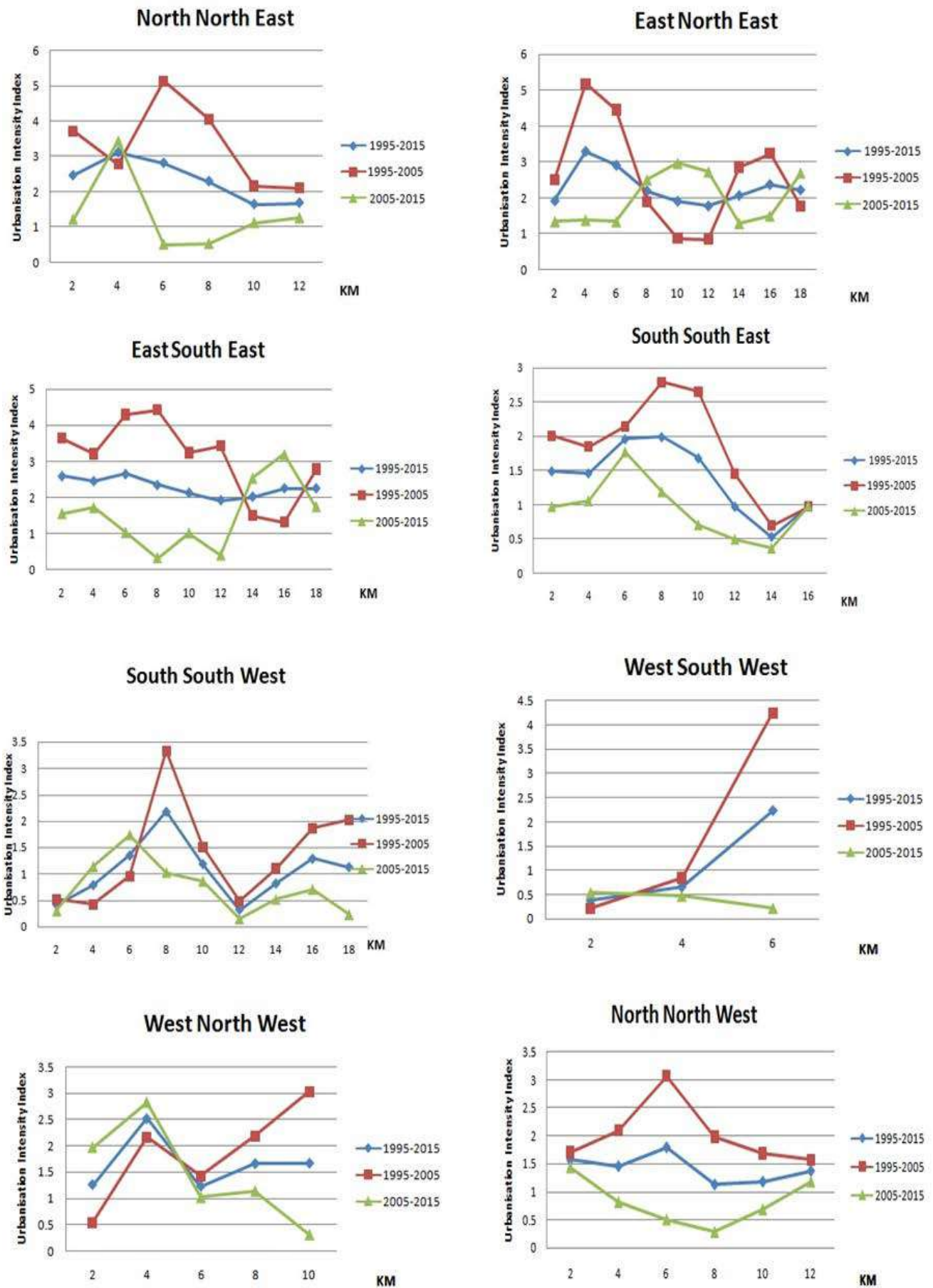


Figure 4. Changes in the UII with distance to the urban centre in different aeolotropic areas over the entire Cochin city region in different time periods from 1995 to 2015

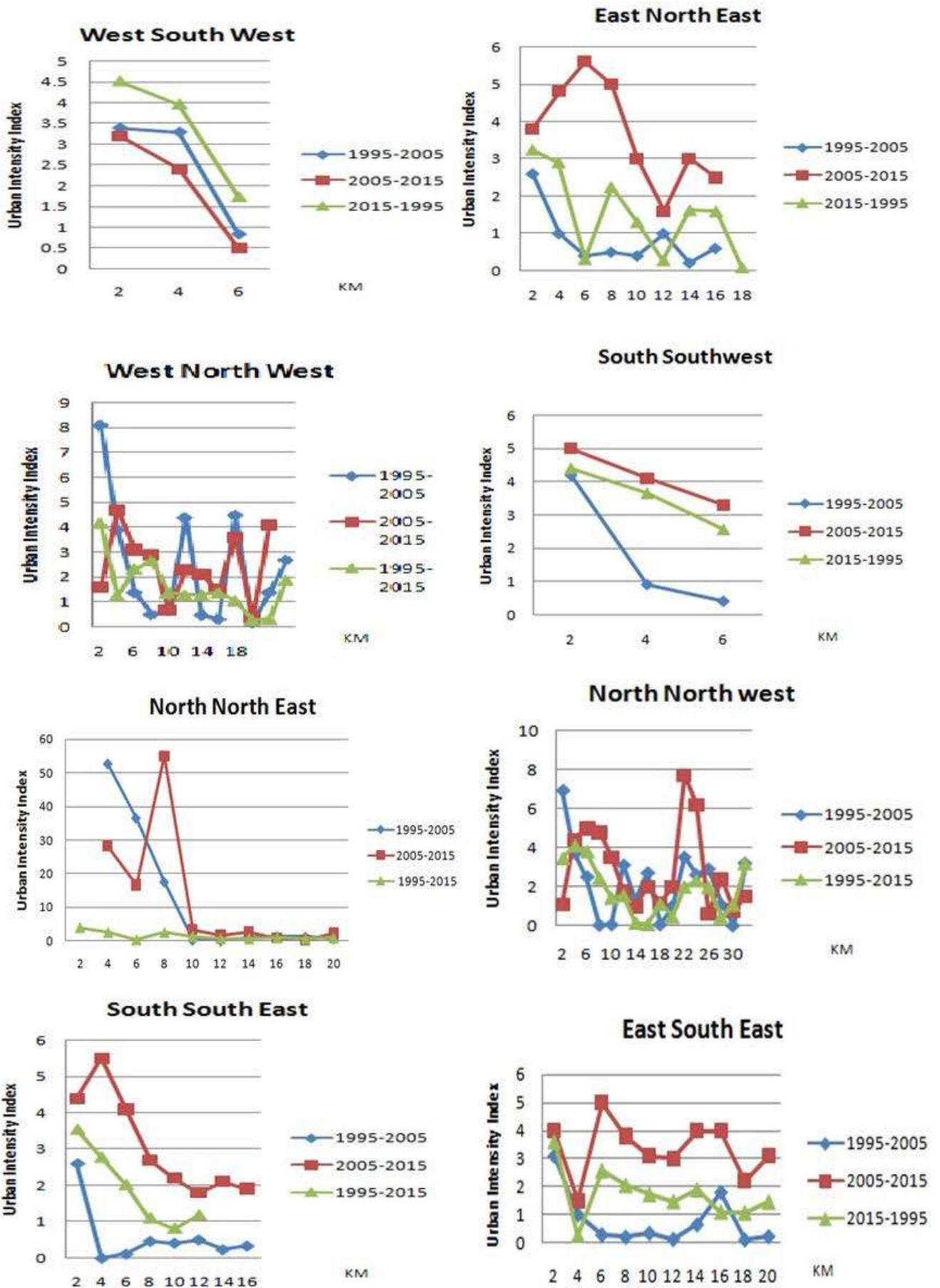


Figure 5. Changes in the UII with distance to the urban centre in different aeolotropic areas over the entire Thiruvananthapuram city region in different time periods from 1995 to 2015

For the period of 2005-2015 Thiruvananthapuram recorded high UPI which is closely interacted with the CBD through various economic activity. NNW, ESE, ENE & NNE spotted its highest UPI in between 4 to 6 Km. These zones are serving the city by providing medical educational and administrative services. Pattom, Medical college, Ulloor, Nemometc are covered by this zone. Between 6 to 10 km NNW, NNE, ESE, ENE and NNE direction UPI is decreasing. Mannanthala, Sreekaryam, peyadetc. Residential population are more concentrated in these zones. The development of IT sector brings more and more investment to the NNE and NNW direction of the city which leads to the further urbanexpansion. The newly developed bypass and well connected road railway network accelerated its expansion. Segmented urbanization is visible in Neyattinkara and Kattakada region, these towns act as a complimentary poles for the cities development and satisfy various functions provided by the city centre. In the NNW slices recorded high UPI at Attingal municipality region.

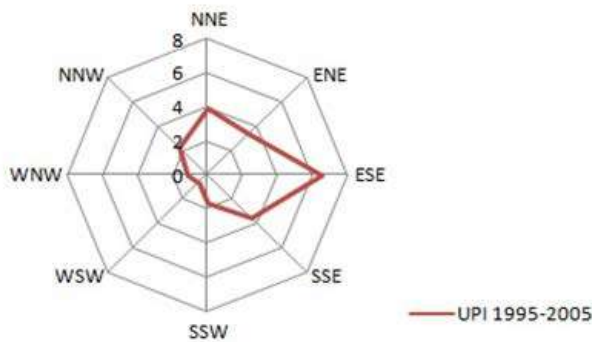
In Cochin UPI curves shows similar increasing trend up to 4km for NNE, ESE, and WNW slices, upto 6 km for SSE and SSW slices, and upto 12 k in ENE slice. Almost all slices reaches its peak value at 6 km as new developmental activities such as IT parks, industries etc. are concentrated in this zone. ENE zone demonstrate a declining trend between 12-6km and after 16 km it shows a rise. ESE slice shows rapid fluctuations in UPI SSW WNW slices shows a similar trend of steady growth and further declining trend, while moving away from the city centre its intensity is declining.

4.3. Spatial Aeolotrophy of UPI Based on Buffer Analysis during 1995-2015

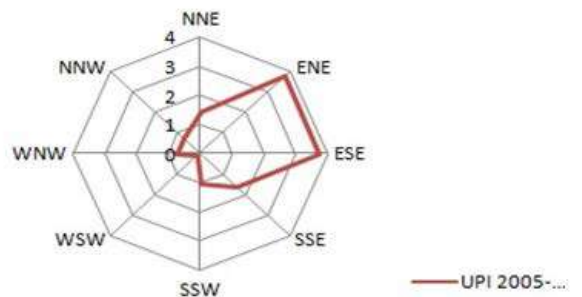
Based on the results of buffer zone analysis, the spatial aeolotrophic characteristics of Cochin and Thiruvananthapuram city region displayed in Figure 6 is summarized as follows. During 1995-2005 urban expansion in Cochin city region concentrated mainly to four directions NNE, ENE, ESE and SSE.

The Urban Expansion is limited in other directions due to the physical constraints (water bodies). Aluva industrial belt is located on the North north eastern direction which accelerates the urban growth to that direction. In ENE direction the developmental activities in Edapally, Thekkumpaddy, Thrikkakara and Kalamassery boosted urban expansion. The well-developed road and rail network as well as the industrial activities in Ambalamugal region hasten the spread of urban area in ESE direction. The apartment villa projects in Maradu and further expansion activities speed up the urbanisation process in SSE direction. For the period of 2005-2015 urban area is spread over ENE and ESE direction. The development of IT parks, industries and commercial activities, institutions and transport development in Kalamassery, Eloor and Thrikkakara led to the spread of urban area in ENE direction. While in ESE direction the urban spread is due to launch of residential buildings in Thripunithura and heritage tourism in Thripunithura and Thiruvankulam, Smart city, Brahmapuram Power Plant, Solid Waste treatment plant in Vavucode-Puthencruz.

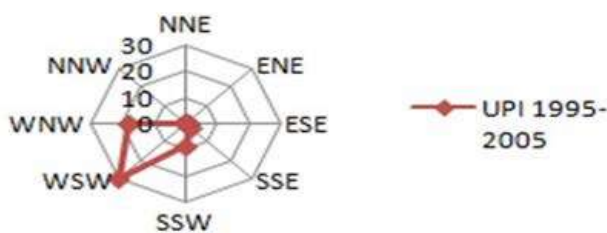
**Cochin City Region
Urban Proportional Index :1995-2005**



**Cochin City Region
Urban Proportional Index :2005-2015**



**Thiruvananthapuram City Region
Urban Proportional Index:1995-2005**



**Thiruvananthapuram City Region
Urban Proportional Index :2005-2015**

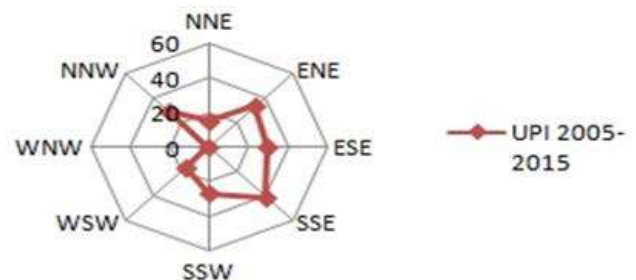


Figure 6. Spatial aeolotropy of Urban Proportional Index based on buffer analysis during 1995-2015

In 1995-2005 the urban expansion of Thiruvananthapuram is less compared to Cochin WNW and WSW direction. The international airport and CBD of the City is in this direction and the city core is much developed in this period compared to other direction. Coastal stretch of the city is located in this direction due to limited area the available peoples densely settled in this region. Manacaud, Chala old market zone, Beemapally, KovalamVeli tourist centres are located in this zone. During 2005-2015 decade the Thiruvananthapuram city expanded much more towards NNW, ENE, ESE, SSW, SSE and NNW. The development in medical, educational and administrative service boosts up this urban growth. Spreading of new IT based industries and more investment in the infrastructural development make the city much more attractive to the migrant population. While analysing the UPI it is clearly evident the city is growing and its spreading toward the outskirt areas.

5. Findings and Conclusion

- Historical legacy played a major role in the growth of these two metropolitan cities of Kerala. Kochi was the princely state came in to existence in 1102 and was under the foreign rule since 18th century. Thiruvananthapuram was under the rule of venad kingdom till 10 century AD. In 1729 Marthandavarma found the princely state of Thiruvithamcore and made Thiruvananthapuram as the capital in 1745. Both these towns cities have trade relation with Indoneasian, French Dutch and Portuguese from the 18th century as famous are the centre spice trade and after the destruction of kodungloor port cochin came into flourished as major trading Hub.
- Post-independence era the economic reforms introduced by central govt of India boosted up the growth of Cochin city region as a financial hub of the state. Thiruvananthapuram retained its position as an administrative centre with an efficient govt machinery in place to conduct the administration of the entire state.
- Geographical position of these two towns influenced its urban growth of these two cities Cochin city geographically located in the central part of Kerala so the accumulation of wealth flowing from central Kerala districts. The Cochin city is well connected with district like Kottayam, Idukki, Alapuzha and Thrissur. Comparing with Cochin Thiruvananthapuram is geographically located in the southwest corner of India having less connectivity to the neighbouring district.
- Kochi metropolitan region having a sustainable economic base through the Private and foreign investment. The service sector flourished after the introduction new economic policy which leads to the rapid commercialization of the city and developed as a financial capital of the state. While comparing with Cochin Thiruvananthapuram city is sustained through Public investment, the city centre is occupied through administrative centres and allied govt headquarters which hindered the private investment. Since 2005 the city gathered its momentum through service sector development, and now the city is growing as a major IT hub having a share of 80% of software export.
- With economic development and population growth, the process of Urbanization of Cochin city had accelerated and the area of urban land increased quickly.
 1. Two distinct phases of urbanisation are discernible
 - a) Phase of Rapid Urbanisation (1995-2005)

This phase is characterised by large scale urban expansion and the urbanisation intensity increases as a whole. There is a rapid expansion in the urban expansion zone. This is because the city centre, Mattancherry, Fort Cochin, Ernakulum regions immediately surrounding the city centre has already developed and urbanised and the areas surrounding it like Maradu, Edapally, Eloor, Kalamassery, Thrikkakara, Thripunithura partly transformed into new urban centres, the urban expansion zone rapidly moves outward with a drastic increase in area. On the outer side of urban expansion zone, the urban expansion intensity decreases.
 - b) Phase of diffusive urbanisation (2005-2015)

Urban expansion has already caused considerable growth of urban area, as well as outward expansion of the urban-suburban transition zone. The extent of urban expansion zone also continues to increase and the top values of UII are at points further from the original urban centres but with lower values. The urbanization is characterized by a transformation from being localized, high intensity and uniform to diffusive, regionalised and complementary. Therefore, UII peak values are lower than those in earlier phase.
 - While analysing the process of Urbanization of Thiruvananthapuram city region it can be categorised into two
 1. The phase of steady urban growth (1995-2005)

This phase was characterised in to urban growth is mainly concentrated in the CBD and the area of influence of the main city centre and a leapfrog urban growth is visible in the municipality areas. The core area is concentrated with administrative activity moving away from the city centre the UII is decreasing because the residential population is also less compared to cochin during this period. Thiruvananthapuram city region is well connected with the suburb so the commuter's population is more within the city. The city growth is mainly based on development tertiary sector
 2. The phase of Uniform urbanization (2005-2015)

In this phase the whole urban intensity of the city is increased compared to the previous decade. Service sector is development and the growth of the city as a major IT hub of the state accelerated the urban growth. More private investments in Tourism, Medical, Educational and research sector leads to the uniform urban growth. The residential population of the also increase in this phase.
 - In cochin the major directional growth was observed to be in the ENE and ESE slice as new developmental activities and projects are undertaking in this zone. Earlier it was also observed in- NNE because of the IT industries and Special Economic Zone in Kalamassery. In Thiruvananthapuram directional growth is towards NNW,ENE,ESEandSSE.The city is growing

towards the suburb panchayaths which lies in between municipalities. The palces such as Nemom, Attingal, Sreekaryam etc having high potential to develop as a satellite town.

- In cochin the urban expansion was observed active for Maradu, Kalamassery and Thripunithura. Residential apartments in Maradu, IT parks at Kalamassery, Heritage tourism and residential buildings in Thripunithura accelerated the urban growth.
- The developmental activities such as Cochin metro rail, Port related activities in Wellington Island, Elamkunnappuzha, Mulavukadu, Njaraikkal, Kadamankkudy, Heritage and tourism activities in Fort Cochin and Mattancherry, Fishing and tourism in Chellanam and Kumbalangi, industrial activities in Vadavucodu-Puthenkurisu, Kalamassery and Thrikkakara, Residential activities in Maradu, Kumbalam, Thripunithura and Thiruvankulam accelerates future trend of urban expansion. While comparing with Cochin Thiruvananthapuram not much explored as an urban centre Kazhakootam, Nedumangad, Neyyatinkara, Varkala and Attingal regions developed as a urban growth poles and has a high potential to develop as a satellite towns. The proposed Vizhinjam Harbour project, Monorail and the new smart city initiatives will boost up the urban growth of the city in future.

6. Conclusion

The Spatio temporal analysis of urban expansion reveals the intensity of urbanization of a particular area and how it influencing landscape of region. This study analyze the directional growth of urban expansion in accordance with its intensity. Through the analysis it reveals that the urban expansion is the product of socio economic and demographic factors individually and sometimes with its association. In the Case of Cochin, the economic demographic & social parameters are going hand in hand. It records a rapid urbanization during 1995-2005 period. Major directional growth of cochin city is mainly towards ENE & ESE direction, which is economically

vibrant region. Thiruvanthapuram is developed as an administrative centre and its urbanization rate is increased during 2005-2015. In both the cities the urbanization is spreading towards the rural stretches were the urban and rural activities mutually correlated. The expansion is towards panchayaths area and which is mainly to meet the residential demand of the city population.

References

- [1] Lipton, Michael. (1977). Why Poor People Stay Poor: Urban Bias in World Development. Temple Smith, London.
- [2] Pacione, M. (2009). Urban Geography: A Global Perspective. Routledge, London.
- [3] United Nations, (2007). Principles and Recommendations for Population and Housing Census, Newyork. Volume 2, 2010, Pages 1139-1156.
- [4] Peter A. Hall, Rosemary C. R. Taylor. (1998). The Potential of Historical Institutionalism: A Response to Hay and Wincott, Political Studies, vol. 46, 5: pp. 958-962.
- [5] Xiaowen Li, (2010). A GIS-based buffer gradient analysis on spatiotemporal dynamics of urban expansion in Shanghai and its major satellite cities. Procedia Environmental Sciences.
- [6] Muchoney, D.M. and Haack, B.N. (1994). Change Detection for Monitoring Forest Defoliation. Photogrammetric Engineering and Remote Sensing, 60, 1243-1251.
- [7] Singh, A. (1989). Digital Change Detection Techniques Using Remotely-Sensed Data. International Journal of Remote Sensing, 10, 989-1003.
- [8] A. Shalaby, R. Tateishi. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt, Elsevier, Applied Geography, 28-41.
- [9] Xiaojun Yang, (2011). Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban, Wiley Blaackwell.
- [10] United Nations. (1993). World Urbanisation Prospects Report, NewYork
- [11] United Nations Environment Programme. (2013). Annual Report.
- [12] Breese G. (1969). The city in the newly delveloped countries, Prentice-Hall
- [13] Chattopadhyay S. (2006). Striving for Sustainability: Environmental Stress and Democratic Initiatives ,Concept Publishing Company, New Delhi.
- [14] T T Sreekumar. (1990). Neither Rural nor Urban: Spatial Formation and Development Process, Economic and Political Weelky Vol 25,1981-1990
- [15] Liu, S., Wu, C., & Shen, H., (2000). A GIS-based model of urban landuse growth in Beijing. Acta Geogr. Sin. 55, 407-416 [in Chinese: English abstract].



ESTIMATION OF THE SCALE PARAMETER OF TRIANGULAR DISTRIBUTION USING ABSOLVED ORDER STATISTICS

ANJANA V*

Department of Statistics, University of Kerala, Thiruvananthapuram, Kerala.

ABSTRACT

In this paper, a new set of ordered random variables called Absolved Order Statistics(AOS) from a scale dependent triangular distribution is considered. The vector of AOS is found to be the minimal sufficient statistic for the triangular distribution with scale parameter σ . Based on AOS, the best linear unbiased estimate $\hat{\sigma}$ of σ along with its variance is explicitly derived. It is found that censoring based on AOS is more realistic and the estimate obtained from it for σ is more efficient than the case of censoring with order statistics. In this study, we also obtained the U-statistic estimator based on AOS for the scale parameter σ of the triangular distribution.

Key words and Phrases: *Triangular Distribution; Order Statistics; Absolved Order Statistics; Minimal Sufficient Statistics; Best Linear Unbiased Estimate; Estimation from Censored Samples; U-statistics.*

*anjana.vpn@gmail.com

1 INTRODUCTION

The scale dependent triangular distribution over $[-\sigma, \sigma]$ is defined by the probability density function (pdf) given by

$$f(x, \sigma) = \frac{1}{\sigma} \left(1 - \frac{|x|}{\sigma}\right), -\sigma < x < \sigma, \sigma > 0, \quad (1.1)$$

where σ is the scale parameter of the distribution. Various extensions of the triangular distribution were discussed by Rene van Dorp and Kotz (2002), Glickman and Xu (2008), and Garg et al. (2009). Johnson and Kotz (1999) described situations wherein triangular distribution is used as an alternative model to the beta distribution. Balakrishnan and Nevzorov (2004) studied some properties of the triangular distribution. Triangular distribution find applications in the Project Evaluation and Review Technique as well as Risk analysis. For details see, Johnson (1997), Fairchild et al. (2016) and Johnson (2002).

Let X be a random variable with pdf (1.1), then $E(X) = 0$ and $V(X) = \frac{\sigma^2}{6}$. Kotz and René Van Dorp (2004) discussed about moments of the triangular distribution in detail and also explained the maximum likelihood estimation procedure for the parameter σ of triangular distribution. The best linear unbiased estimation in triangular distribution was carried out using order statistics in Samuel and Thomas (2003). The maximum likelihood (ML) method of estimation of σ in (1.1) is somewhat not so trivial as the likelihood function is not differentiable at σ . The form of the density (1.1) and the likelihood resulting from it fails to provide the ML estimate of σ explicitly. Several authors though tried to obtain the ML estimate, they end up with procedures of obtaining the estimate involved with a software that do not contain detailed descriptions of their procedure. In this respect the estimate provided by Samuel and Thomas (2003) attracts some importance at least in small sample cases.

Thomas and Anjana (2021) have considered the family \mathcal{F}_1 of all absolutely continuous distributions which are symmetrically distributed about zero and defined a new variety of ordered random variables and is given below

Definition 1.1. (Thomas and Anjana, 2021). Suppose X_1, X_2, \dots, X_n is a random sample of size n drawn from a distribution with pdf $f(x)$ such that $f(x) \in \mathcal{F}_1$. If we take the absolute values of the observations and order them in the increasing order of magnitude as $X_{(1:n)} \leq X_{(2:n)} \leq \dots \leq X_{(n:n)}$, then we say that $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the absolute order statistics (AOS) of the given sample.

Thomas and Anjana (2021) further proved the following theorem on the minimal sufficiency of the statistics introduced in the above definition.

Theorem 1.1. (see, Theorem 2.1 of Thomas and Anjana, 2021). Suppose $\tilde{X} = (X_1, X_2, \dots, X_n)$ is a vector of observations of a random sample of size n drawn from a distribution with pdf $f(x) \in \mathcal{F}_1$. Let $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ be a statistic based on the AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ constructed from the observations X_1, X_2, \dots, X_n . Then T is minimal sufficient for the family \mathcal{F}_1 .

The vector of order statistics $U = (X_{1:n}, X_{2:n}, \dots, X_{n:n})'$ of a random sample X_1, X_2, \dots, X_n of size n arising from a distribution belonging to the family \mathcal{F} of absolutely continuous distributions was proved as complete and minimal sufficient statistic for \mathcal{F} . As $\mathcal{F}_1 \subset \mathcal{F}$, no thought is seen went through the minds of authors about the improvement of minimal sufficient statistic for the sub-class \mathcal{F}_1 of \mathcal{F} . The Theorem 1.1 proved by Thomas and Anjana (2021) provided the improvement of the minimal sufficient statistic for \mathcal{F}_1 . Clearly the pdf (1.1) belongs to \mathcal{F}_1 and hence the BLUE based on AOS should have an advantage over the BLUE based on order statistics in the estimation of the parameter σ involved in (1.1).

In Section 2, we derive the distribution theory of AOS arising from (1.1) and proved that vector of AOS is minimal sufficient for the family of triangular distributions as well. In Section 3, we have discussed about the estimate of σ in (1.1) by the BLUE based on AOS. The efficiency of this estimate relative to the usual BLUE based on order statistics for sample sizes $n = 2(1)20$ is computed and analyzed. Since this estimate of σ for triangular distribution is not perceived well, another type of estimator called U-statistic defined from BLUE based on AOS as kernel is attempted in section 5. It is to be noted that U-statistics are unbiased consistent estimators and they are normally distributed as $n \rightarrow \infty$.

2 MINIMAL SUFFICIENT STATISTIC FOR THE TRIANGULAR DISTRIBUTION

Let X_1, X_2, \dots, X_n be a random sample of size n drawn from the triangular distribution with pdf as given in (1.1). Suppose $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of the sample. Then the distribution theory of AOS is given in the following theorem.

Theorem 2.1. *Suppose $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of a random sample of size n drawn from the distribution with pdf $f(x, \sigma)$ which is as given in (1.1). Let $Z_{1:n}, Z_{2:n}, \dots, Z_{n:n}$ be the order statistics of a random sample of size n arising from the folded distribution of (1.1) with density $g(z, \sigma) = 2f(z, \sigma), 0 \leq z < \sigma$. Then*

$$(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}) \stackrel{d}{=} (Z_{1:n}, Z_{2:n}, \dots, Z_{n:n}),$$

where $X \stackrel{d}{=} Z$ is the usual notation representing the identically distributed property between two random variables X and Z .

Proof. Let X follow triangular distribution with pdf (1.1). Let $Z = |X|$, then Z with pdf $g(z, \sigma) = 2f(z, \sigma), 0 \leq z < \sigma$, follows half-triangular distribution. Since $X_{(1:n)} \leq X_{(2:n)} \leq \dots \leq X_{(n:n)}$ are the order statistics of the random sample $|X_1|, |X_2|, \dots, |X_n|$. We have $(X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ is distributed identically as the vector of order statistics of a random sample of size n arising from half-triangular distribution. Hence the theorem. \square

Now, let us discuss about a minimal sufficient statistic for the triangular family of distributions as defined by the pdf (1.1). From theorem 1.1 due to Thomas and Anjana (2021), it is observed that the statistic $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ formed from the AOS of a random sample of size n is the minimal sufficient statistic for the family \mathcal{F}_1 of distributions which are all symmetric about zero. If we consider a specific member of \mathcal{F}_1 with pdf $f(x)$, then some times it is possible to get a dimensional reduction in the minimal sufficient statistic for $f(x)$. For example, some distributions belonging to the exponential family (such as normal distribution) which are distributed symmetrically about zero, we get a minimal sufficient statistic

which is different from $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ and whose dimension is less than n . The essential point is that by merely observing triangular distribution as defined in (1.1) belongs to \mathcal{F}_1 , we cannot claim that $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ is also minimal sufficient for triangular distribution. Clearly, triangular distribution is not a member of exponential family and similarly we can verify that triangular distribution is not a member to any sub-family of the class of all continuous distributions for which an improved version of minimal sufficient statistic exists in terms of reduced dimension than $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$. Observing the above arguments about the irreducible nature of the sufficient statistic than T , we can prove the following theorem using basic principles.

Theorem 2.2. *Suppose $\underline{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n drawn from the triangular distribution with pdf $f(x, \sigma)$ as given in (1.1). Let $T = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})$ be the vector of AOS constructed from X_1, X_2, \dots, X_n . Then T is minimal sufficient for $f(x, \sigma)$.*

The proof of this theorem is not given here similar to the proof given by Thomas and Anjana (2021). Hence omitted.

It is well known in statistical inference that any inference procedure developed based on minimal sufficient statistic excel in performance than those based on other statistics. Thus at this stage it is destined to observe that the Best Linear Unbiased Estimate (BLUE) of σ based on AOS is better than that based on the classical order statistics. In the next section this aspect is discussed more.

3 BEST LINEAR UNBIASED ESTIMATION OF THE SCALE PARAMETER BASED ON AOS

There is extensive literature available about the problem of estimation of the location and scale parameters of a distribution by order statistics. For details see, David and Nagaraja (2003), Balakrishnan and Cohen (1991). For modification of the BLUE for symmetric populations see, Thomas (1990). In this section by a similar manner we apply AOS to estimate the scale parameter σ of triangular distribution. The

standard form of triangular distribution as given in (1.1) has the pdf

$$f(x) = \left(1 - |x|\right), -1 < x < 1. \tag{3.1}$$

The expressions for the single and product moments of order statistics arising from the two parameter triangular distribution along with best linear unbiased estimators of the location and scale parameters based on order statistics were obtained in Samuel and Thomas (2003). From (1.1) the pdf of the half-triangular distribution can be written as

$$g(x, \sigma) = \frac{2}{\sigma} \left(1 - \frac{x}{\sigma}\right), 0 \leq x \leq \sigma, \sigma > 0. \tag{3.2}$$

The pdf of the corresponding standard form of half-triangular distribution is

$$g(x, 1) = 2(1 - x), 0 \leq x \leq 1. \tag{3.3}$$

In the following theorem, the expression for the BLUE and its variance for the scale parameter σ of the triangular distribution based on AOS is given.

Theorem 3.1. *Let $\underline{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})'$ be the first n AOS of a random sample of size n drawn from the triangular distribution defined in (1.1). Suppose $\underline{Y} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})'$ is the vector of order statistics arising from the standard half-triangular distribution defined by the pdf (3.3). Let $E(\underline{Y}) = \underline{\alpha} = (\alpha_{1:n}, \alpha_{2:n}, \dots, \alpha_{n:n})'$ and let the dispersion matrix of \underline{Y} be given by $D(\underline{Y}) = A = ((\alpha_{i,j:n}))$, where $\alpha_{i,j:n} = Cov(Y_{i:n}, Y_{j:n}), i, j = 1, 2, \dots, n$, for $i \neq j$ and $\alpha_{i,i:n} = Var(Y_{i:n})$ for $i = 1, 2, \dots, n$. Then the BLUE $\hat{\sigma}$ of σ based on AOS is given by*

$$\hat{\sigma} = (\underline{\alpha}' \mathbf{A}^{-1} \underline{\alpha})^{-1} \underline{\alpha}' \mathbf{A}^{-1} \underline{X} \tag{3.4}$$

and its variance is given by

$$Var(\hat{\sigma}) = (\underline{\alpha}' \mathbf{A}^{-1} \underline{\alpha})^{-1} \sigma^2. \tag{3.5}$$

Proof. Let $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ be the AOS of a random sample of size n drawn from the triangular distribution with pdf (1.1). Then $(\frac{X_{(1:n)}}{\sigma}, \frac{X_{(2:n)}}{\sigma}, \dots, \frac{X_{(n:n)}}{\sigma})' \stackrel{d}{=} (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})'$ where $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$ are the order statistics of a random sample of size n drawn from the standard half-triangular distribution with pdf (3.3).

Clearly the means, variances and covariances of all order statistics involved in the right side vector of above distributional identity exists finitely and they are independent of σ . Thus for $\underset{\sim}{X} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)})'$ and $\underset{\sim}{Y} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n:n})'$ we can write

$$E(\underset{\sim}{X}) = \sigma E(\underset{\sim}{Y}) = \sigma \underset{\sim}{\alpha} \quad (3.6)$$

and

$$D(\underset{\sim}{X}) = \sigma^2 D(\underset{\sim}{Y}) = \sigma^2 A. \quad (3.7)$$

Then (3.6) and (3.7) together form a generalized Gauss-Markov setup and hence by Gauss-Markov theorem the BLUE of σ is given by

$$\hat{\sigma} = (\underset{\sim}{\alpha}' \mathbf{A}^{-1} \underset{\sim}{\alpha})^{-1} \underset{\sim}{\alpha}' \mathbf{A}^{-1} \underset{\sim}{X} \quad (3.8)$$

and

$$Var(\hat{\sigma}) = (\underset{\sim}{\alpha}' \mathbf{A}^{-1} \underset{\sim}{\alpha})^{-1} \sigma^2. \quad (3.9)$$

This proves the theorem. \square

The linear estimate $\hat{\sigma}$ of σ as given in (3.8) may also be written as

$$\hat{\sigma} = \sum_{i=1}^n c_{i,n} X_{(i:n)}, \quad (3.10)$$

where $c_{i,n}$, $i = 1, 2, \dots, n$ are appropriate constants.

In (3.10), it is strange to note that $X_{(i:n)}$, $i = 1, 2, \dots, n$ are the AOS arising from the triangular distribution defined in (1.1), while $c_{i,n}$, $i = 1, 2, \dots, n$ are the coefficients of BLUE of σ based on the order statistics of a random sample of size n arising from the half-triangular distribution defined by the pdf (3.2) (a consequence of Theorem 2.1). This makes us to conclude that the BLUE of σ based on order statistics of a sample of size n arising from half-triangular distribution as defined by the pdf (3.3) and its variance is available, then the BLUE of σ based on the AOS of a sample of size n arising from the triangular distribution as defined by the pdf (1.1) and its variance can be obtained without any direct evaluation of means, variances and covariances of those AOS.

The moment relations for order statistics arising from general distributions which are symmetrically distributed about zero such as those described in Arnold et al. (1992), David and Nagaraja (2003) and Thomas and Samuel (1996) are useful for the easy evaluation of moments of order statistics from symmetric distributions, which can also apply to triangular distribution as such. However, application of Theorem 2.1 helps us to deal only with the evaluation of moments of order statistics arising from half-triangular distribution for using those values to the development of inference procedures on the scale parameter σ of triangular distribution. Moments of order statistics arising from standard half-triangular distribution is not seen discussed in the available literature. Hence, we have used Mathematica software to evaluate the means, variances and covariances of all order statistics $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$ for $n = 2(1)20$ arising from the standard half-triangular distribution defined by the pdf (3.3). Using those values we have determined the coefficients $c_{i,n}$ of the AOS $X_{(i:n)}$ in the estimate $\hat{\sigma}$ of σ as given in (3.10) for $i = 1, 2, \dots, n; n = 2(1)20$ and those values are presented in Table 1. We have obtained further $\sigma^{-2}Var(\hat{\sigma})$ for $n = 2(1)20$ and those values are given in Table 2. Samuel and Thomas (2003) obtained the BLUE σ^* of σ based on order statistics of a random sample of size n from the triangular distribution with pdf (1.1). We have tabulated $Var(\sigma^*)$ for $n = 2(1)20$ and is given in Table 2 with an objective of comparing the performance of the new estimator $\hat{\sigma}$ proposed in this paper with the already available estimator σ^* .

Table 1: Coefficient $c_{i,n}$ of $X_{(i:n)}$ involved in $\hat{\sigma} = \sum_{i=1}^n c_{i,n} X_{(i:n)}$ as an estimate of σ of triangular distribution for $n=2(1)20$.

n	Coefficient $c_{i,n}$ of $X_{(i:n)}$ involved in $\hat{\sigma} = \sum_{i=1}^n c_{i,n} X_{(i:n)}$																				
	$X_{(1:n)}$	$X_{(2:n)}$	$X_{(3:n)}$	$X_{(4:n)}$	$X_{(5:n)}$	$X_{(6:n)}$	$X_{(7:n)}$	$X_{(8:n)}$	$X_{(9:n)}$	$X_{(10:n)}$	$X_{(11:n)}$	$X_{(12:n)}$	$X_{(13:n)}$	$X_{(14:n)}$	$X_{(15:n)}$	$X_{(16:n)}$	$X_{(17:n)}$	$X_{(18:n)}$	$X_{(19:n)}$	$X_{(20:n)}$	
2	0.3333	2.0000																			
3	0.1818	0.2727	1.6364																		
4	0.1200	0.1600	0.2400	1.4400																	
5	0.0876	0.1095	0.1460	0.2190	1.3139																
6	0.0681	0.0816	0.1021	0.1361	0.204	1.2245															
7	0.0551	0.0643	0.0771	0.0964	0.1286	0.1928	1.157														
8	0.0459	0.0526	0.0613	0.0736	0.0919	0.1227	0.1839	1.1038													
9	0.0393	0.0442	0.0505	0.0589	0.0706	0.0884	0.1178	0.1767	1.0605												
10	0.0341	0.0379	0.0427	0.0488	0.0569	0.0683	0.0854	0.1138	0.1707	1.0243											
11	0.0301	0.0331	0.0367	0.0414	0.0473	0.0552	0.0662	0.0828	0.1104	0.1656	0.9934										
12	0.0269	0.0293	0.0322	0.0358	0.0403	0.0461	0.0537	0.0645	0.0806	0.1074	0.1612	0.9667									
13	0.0241	0.0264	0.0285	0.0315	0.0349	0.0393	0.0449	0.0524	0.0629	0.0786	0.1048	0.1572	0.9434								
14	0.0219	0.0237	0.0256	0.0279	0.0307	0.0342	0.0384	0.0439	0.0513	0.0615	0.0769	0.1025	0.1538	0.9226							
15	0.0201	0.0215	0.0232	0.0251	0.0274	0.0301	0.0335	0.0377	0.043	0.0503	0.0603	0.0755	0.1005	0.1507	0.9041						
16	0.0185	0.0197	0.0211	0.0228	0.0247	0.0269	0.0296	0.0329	0.0368	0.0424	0.0493	0.0592	0.0739	0.0986	0.1479	0.8874					
17	0.0171	0.0182	0.0194	0.0208	0.0224	0.0243	0.0264	0.0291	0.0323	0.0363	0.0415	0.0485	0.0581	0.0727	0.0969	0.1454	0.8722				
18	0.0159	0.0169	0.0179	0.0191	0.0204	0.022	0.0239	0.026	0.0286	0.0318	0.0358	0.0409	0.0477	0.0572	0.0715	0.0954	0.1431	0.8584			
19	0.0148	0.0157	0.0166	0.0175	0.0189	0.0201	0.0217	0.0235	0.0256	0.0282	0.0314	0.0352	0.0403	0.0469	0.0564	0.0705	0.0939	0.1409	0.8456		
20	0.0138	0.0148	0.0154	0.0164	0.0174	0.0185	0.0199	0.02195	0.0271	0.0234	0.0279	0.0308	0.0347	0.0397	0.0463	0.0556	0.0695	0.0927	0.1389	0.8339	

Clearly, as $\hat{\sigma}$ depends on the observations which are the components of minimal sufficient statistic, the estimate $\hat{\sigma}$ is likely to be more efficient than the estimate σ^* . Thus our proposed new estimator $\hat{\sigma}$ of σ based on AOS is more preferable. To compare the performance of our estimator $\hat{\sigma}$, the relative efficiency $e(\hat{\sigma}/\sigma^*)$ is defined as $e(\hat{\sigma}/\sigma^*) = \frac{Var(\sigma^*)}{Var(\hat{\sigma})}$. The relative efficiencies for $n = 2(1)20$ are calculated and are presented in Table 2. From Table 2, it may be concluded that the estimate $\hat{\sigma}$ based on AOS is remarkably better than σ^* based on order statistics. The gain in efficiency observed in $\hat{\sigma}$ when compared with σ^* ranges from 25% to 138%.

4 ESTIMATION OF THE SCALE PARAMETER FROM CENSORED SAMPLES

When an outlier occurs in a sample drawn from a distribution which is symmetric about zero, it must be far off from zero either in the positive side or negative side. It is quite curious to know that unlike order statistics of the data, AOS capture that far off observation uniquely as the largest absolved order statistic $X_{(n:n)}$. But while working with order statistics this uniqueness is not materialized, as the farthest observation from zero may be either the smallest order statistic $X_{1:n}$ or the largest order statistic $X_{n:n}$. So to eliminate the effect of a suspected outlier from the sample with ordered data, theoretically a double censoring with one observation in the left ($X_{1:n}$) and one observation in the right ($X_{n:n}$) is required. Though for the triangular distribution also the above descriptions applies if we suspect more than two outliers in the data, then we have to generalize the estimate of σ by modifying the censoring scheme appropriately. In particular, if we have the reason to believe that there are k outlying observations in the data, then we go for right censoring of k of the AOS (those corresponding to the k observations in the data which lie most distantly from zero than others) and then estimate σ by using the available AOS $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)}$ where k is any positive integer such that $1 \leq k \leq n - 2$. The estimation procedure for σ in the censoring scheme then follows from the theorem given below.

Table 2: Variances of (i) BLUE $\hat{\sigma}$ based on AOS (ii) BLUE σ^* based on order statistics of the scale parameter σ of triangular distribution and the relative efficiency : $e(\hat{\sigma}/\sigma^*)$ for $n=2(1)20$.

n	$\sigma^{-2}Var(\hat{\sigma})$	$\sigma^{-2}Var(\sigma^*)$	$e(\hat{\sigma}/\sigma^*)$
2	0.2222	0.5306	2.3879
3	0.1364	0.2415	1.7705
4	0.0960	0.1511	1.5739
5	0.0730	0.1080	1.4795
6	0.0583	0.0830	1.4237
7	0.0482	0.0668	1.3859
8	0.0409	0.0555	1.3570
9	0.0354	0.0473	1.3362
10	0.0311	0.0409	1.3151
11	0.0276	0.0361	1.3080
12	0.0248	0.0321	1.2944
13	0.0225	0.0289	1.2845
14	0.0205	0.0262	1.2781
15	0.0188	0.0239	1.2713
16	0.0173	0.0220	1.2644
17	0.0162	0.0203	1.2531
18	0.0151	0.0188	1.2450
19	0.0141	0.0176	1.2482
20	0.0132	0.0165	1.2500

Theorem 4.1. *Suppose $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of a random sample of size n drawn from the triangular distribution with pdf defined in (1.1). Let $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$ be the order statistics of a random sample of size n drawn from the standard half-triangular distribution given in (3.3). Suppose k observations corresponding to those with largest k absolute values are censored so that the vector of the remaining AOS is $X_{\sim n-k} = (X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)})'$. Define $Y_{\sim n-k} = (Y_{1:n}, Y_{2:n}, \dots, Y_{n-k:n})'$, $E(Y_{\sim n-k}) = \alpha_{\sim n-k} = (\alpha_{1:n}, \alpha_{2:n}, \dots, \alpha_{n-k:n})'$ and let the dispersion matrix of $Y_{\sim n-k}$ is denoted by A_{n-k} . In this case, we write $E(X_{\sim n-k}) = \alpha_{\sim n-k} \sigma$, $D(X_{\sim n-k}) = A_{n-k} \sigma^2$. Then the BLUE $\hat{\sigma}_{k,n}$ of σ based on the censored AOS is given by*

$$\hat{\sigma}_{k,n} = (\alpha'_{\sim n-k} \mathbf{A}_{n-k}^{-1} \alpha_{\sim n-k})^{-1} \alpha'_{\sim n-k} \mathbf{A}_{n-k}^{-1} X_{\sim n-k}. \quad (4.1)$$

The variance of $\hat{\sigma}_{k,n}$ is given by

$$\text{Var}(\hat{\sigma}_{k,n}) = (\alpha'_{\sim n-k} \mathbf{A}_{n-k}^{-1} \alpha_{\sim n-k})^{-1} \sigma^2. \quad (4.2)$$

Proof. The proof of the above theorem follows easily by the application of Gauss-Markov theorem. \square

Remark 4.1. *One can write (4.1) as a linear function of $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n-k:n)}$ as*

$$\hat{\sigma}_{k,n} = \sum_{i=1}^{n-k} c_{i,n}^{(k)} X_{(i:n)}, \quad (4.3)$$

where $c_{i,n}^{(k)}$, $i = 1, 2, \dots, n - k$ are appropriate constants.

Table 3: Coefficient $c_{i,10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,10} = \sum_{i=1}^{10-k} c_{i,10}^{(k)} X_{(i:10)}$, $Var(\hat{\sigma}_{k,10})$ and efficiency $e(\hat{\sigma}_{k,10}/\hat{\sigma})$ of $\hat{\sigma}_{k,10}$ relative to $\hat{\sigma}$ for the scale parameter σ of the triangular distribution.

Statistic used	Coefficient $c_{i,10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,10} = \sum_{i=1}^{10-k} c_{i,10}^{(k)} X_{(i:10)}$ given along columns.									
	$\hat{\sigma}$ ($k=0$)	$\hat{\sigma}_{1,10}$ ($k=1$)	$\hat{\sigma}_{2,10}$ ($k=2$)	$\hat{\sigma}_{3,10}$ ($k=3$)	$\hat{\sigma}_{4,10}$ ($k=4$)	$\hat{\sigma}_{5,10}$ ($k=5$)	$\hat{\sigma}_{6,10}$ ($k=6$)	$\hat{\sigma}_{7,10}$ ($k=7$)	$\hat{\sigma}_{8,10}$ ($k=8$)	
$X_{(1:10)}$	0.0341	0.0518	0.0699	0.0913	0.1183	0.1549	0.2088	0.2975	0.4737	
$X_{(2:10)}$	0.0379	0.0576	0.0778	0.1014	0.1314	0.1721	0.2320	0.3306	10.000	
$X_{(3:10)}$	0.0427	0.0648	0.0875	0.1141	0.1478	0.1936	0.2610	6.3223		
$X_{(4:10)}$	0.0488	0.0741	0.0999	0.1304	0.1689	0.2213	4.4739			
$X_{(5:10)}$	0.0569	0.0864	0.1166	0.1521	0.1971	3.3559				
$X_{(6:10)}$	0.0683	0.1037	0.1399	0.1826	2.6016					
$X_{(7:10)}$	0.0854	0.1296	0.1749	2.0536						
$X_{(8:10)}$	0.1138	0.1728	1.6329							
$X_{(9:10)}$	0.1707	1.2960								
$X_{(10:10)}$	1.0243									
$\sigma^{-2}Var(\hat{\sigma}_{k,n})$	0.0310	0.0471	0.0636	0.0830	0.1075	0.1408	0.1898	0.2705	0.4306	
$e(\hat{\sigma}_{k,n}/\hat{\sigma})$	-	0.6586	0.4879	0.3741	0.2887	0.2204	0.1635	0.1148	0.0721	

The method of estimation of σ by $\hat{\sigma}_{k,n}$ as described in Theorem 4.1 using censored AOS arising from the triangular distribution is attempted for a sample of size 10 and for each of $k = 1, 2, 3, \dots, 8$. For each $k = 1, 2, 3, \dots, 8$, we have computed the numerical value of the coefficients $c_{i,10}^{(k)}$ of $X_{(i:10)}$ involved in $\hat{\sigma}_{k,10}$ for $i = 1, 2, \dots, 10-k$ and $\sigma^{-2}Var(\hat{\sigma}_{k,10})$ and these computed values are given in Table 3. The relative efficiency $e(\hat{\sigma}_{k,10}/\hat{\sigma}) = \frac{Var(\hat{\sigma})}{Var(\hat{\sigma}_{k,10})}$ of $\hat{\sigma}_{k,10}$ when compared with the BLUE $\hat{\sigma}$ as derived in (3.10) is again computed for each $k = 1, 2, 3, \dots, 8$ and the computed values are also presented in Table 3. Note that when $k = 0$, the estimate $\hat{\sigma}_{0,10}$ is same as $\hat{\sigma}$ as given in (3.10) for $n = 10$.

From Table 3, we observed that initially for $k = 1$, there is not much reduction noticed on the relative efficiency keeping the relative efficiency more than 65%. However the reduction noticed in the relative efficiencies become little more but changes at a sluggish rate when more number of extreme absolute order statistics are censored. For example, if $n = 10$ and $k = 8$ then in the estimator $\hat{\sigma}_{8,10}$ altogether eight AOS are censored and hence it utilizes only two AOS, $X_{(1:10)}$ and $X_{(2:10)}$ for estimating σ . From these two AOS, the efficiency observed in $\hat{\sigma}_{8,10}$ relative to the estimator $\hat{\sigma}$ (in which 8 out of 10 AOS are involved) is more than 7%. This makes to comment that AOS based estimators for the scale parameter σ of triangular distribution appears to be robust.

5 U-STATISTICS AS ESTIMATOR FOR THE SCALE PARAMETER

Sreekumar and Thomas (2007) have introduced the technique of constructing U-statistic estimators for the location and scale parameters of a distribution by taking BLUE based on order statistics of those parameters as kernels. Sreekumar and Thomas (2007) have also narrated a technique of obtaining the variance of those U-statistics and illustrated their method to estimate the parameters of log-gamma distribution. For applications of U-statistics based on BLUE's based on order statistic in estimating the location and scale parameters of distributions see also, Sreeku-

mar and Thomas (2008), Thomas and Baiju (2012, 2015), Thomas and Priya (2015, 2016), Sreekumar and Thomas (2008). If \mathcal{F}_1 is the family of distributions which are all symmetrically distributed about zero, Thomas and Anjana (2021) explained that in order to deal with the estimation of σ based on AOS arising from $f(x, \sigma) \in \mathcal{F}_\theta^{(1)}$, it is enough to deal with the estimation of σ based on order statistics arising from the folded distribution with pdf $g(z, \sigma) = \frac{2}{\sigma} f_0(\frac{z}{\sigma}), 0 < z < \infty$. Then to obtain the estimate of σ based on AOS arising out of a random sample of size n from $f(x, \sigma)$, we attempt for the corresponding estimate of σ based on order statistics of a random sample of size n arising from the pdf $g(z, \sigma)$ and replace in it each order statistic by the correspondingly ordered AOS. The variance expression for the estimates remains the same for both cases. As a result of these logical arguments, Thomas and Anjana (2021) derived a U-statistic using BLUE based on AOS as kernel for σ involved in $f(x, \sigma)$ and also showed that it is enough to derive a U-statistic using BLUE based on order statistics as kernel for σ involved in $g(z, \sigma)$ and finally replace each order statistic in the estimate by the correspondingly ordered AOS. Now the method of constructing the U-statistic for σ involved in triangular distribution is described here using a BLUE based on absolved order statistics.

Let X_1, X_2, \dots, X_n be a random sample of size n drawn from triangular distribution, defined in (1.1) so that the absolute values of the observations $|X_1|, |X_2|, \dots, |X_n|$ may be considered as a random sample of size n drawn from half-triangular distribution defined in (3.2). Let Z_1, Z_2, \dots, Z_m be an initial random sample of size m drawn from the half- triangular distribution defined in (3.2). Let $\tilde{Z} = (Z_{1:m}, Z_{2:m}, \dots, Z_{m:m})'$ be the vector of order statistics obtained from Z_1, Z_2, \dots, Z_m . If $\tilde{V} = (V_{1:m}, V_{2:m}, \dots, V_{m:m})'$ is considered as the vector of order statistics of a random sample of size m drawn from the pdf given by (3.3), then the distribution of these order statistics are independent of σ .

Let $E(\tilde{V}) = \tilde{\beta} = (\beta_{1:m}, \beta_{2:m}, \dots, \beta_{m:m})'$ be the vector of expectation of \tilde{V} and let the dispersion matrix of \tilde{V} be denoted by \mathbf{G} . Then from David and Nagaraja (2003), the BLUE of σ involved in (3.2) based on order statistics is given by

$$\hat{\sigma} = (\tilde{\beta}' \mathbf{G}^{-1} \tilde{\beta})^{-1} \tilde{\beta}' \mathbf{G}^{-1} \tilde{Z} \quad (5.1)$$

Table 4: Variances components $\xi_c^{(m)}$ for $c = 1(1)m$ and $m = 2(1)5$ for triangular distribution.

m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$	m	c	$\sigma^{-2}\xi_c^{(m)}$
2	1	0.1009	4	1	0.0209	5	1	0.0125
	2	0.2222		2	0.0435		2	0.0260
	1	0.0402		3	0.0683		3	0.0403
3	2	0.0851		4	0.0960		4	0.0559
	3	0.1364			5		0.0730	

and variance of $\hat{\sigma}$ is given by

$$Var(\hat{\sigma}) = (\beta' \underset{\sim}{\mathbf{G}}^{-1} \underset{\sim}{\beta})^{-1} \sigma^2. \tag{5.2}$$

We may write $\hat{\sigma}$ also as

$$h(Z_1, Z_2, \dots, Z_m) = c_{1:m}Z_{1:m} + c_{2:m}Z_{2:m} + \dots + c_{m:m}Z_{m:m}, \tag{5.3}$$

where $c_{1:m}, c_{2:m}, \dots, c_{m:m}$ are constants which are determined from (5.1). Now from Sreekumar and Thomas (2007), we can easily write the U-statistic for a random sample Z_1, Z_2, \dots, Z_n ($n > m$) drawn from the distribution with pdf (3.2) based on the kernel (5.3) as

$$U_n^{(m)} = \frac{1}{\binom{n}{m}} \sum_{r=1}^n \left[\sum_{j=1}^m \binom{n-r}{m-j} \binom{r-1}{j-1} c_{j:m} \right] Z_{r:n}, \tag{5.4}$$

where we define for non-negative integers p and q, $\binom{p}{q} = 0$ for $p < q$. If we write

$$\xi_c^{(m)} = Cov[h(Z_1, Z_2, \dots, Z_c, Z_{c+1}, \dots, Z_m), h(Z_1, Z_2, \dots, Z_c, Z_{m+1}, \dots, Z_{2m-c})], \tag{5.5}$$

for, $c = 1, 2, \dots, m$, then the $Var(U_n^{(m)})$ is given by

$$Var[U_n^{(m)}] = \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \xi_c^{(m)}. \tag{5.6}$$

Clearly $\xi_m^{(m)} = \text{Var}(h(Z_1, Z_2, \dots, Z_m))$ and is given by (5.2).

The components $\xi_c^{(m)}$ for $c = 1, 2, \dots, m - 1$ involved is $\text{Var}[U_n^{(m)}]$ though look very simple, obtaining those values exactly is somewhat computationally difficult with respect to a given kernel. However Thomas and Anjana (2021) have given a methodology to obtain those values and the version of that methodology to the present problem is given below.

If we put $n = m + k$ in (5.6), then we obtain

$$\text{Var}(U_{m+k}^{(m)}) = \frac{1}{\binom{m+k}{m}} \left\{ \binom{m}{m-k} \binom{k}{k} \xi_{m-k}^{(m)} + \binom{m}{m-k+1} \binom{k}{k-1} \xi_{m-k+1}^{(m)} + \dots + \binom{m}{m} \binom{k}{0} \xi_m^{(m)} \right\}, \quad (5.7)$$

for, $k = 1, 2, \dots, m - 1$.

On putting $n = m + k$ in (5.4) we obtain

$$U_{m+k}^{(m)} = \frac{1}{\binom{m+k}{m}} \left\{ \left[\sum_{j=1}^m \binom{m-k-1}{m-j} \binom{0}{j-1} c_{j:m} \right] Z_{1:m+k} + \left[\sum_{j=1}^m \binom{m+k-2}{m-j} \binom{1}{j-1} c_{j:m} \right] Z_{2:m+k} + \dots + \left[\sum_{j=1}^m \binom{0}{m-j} \binom{m+k-1}{j-1} c_{j:m} \right] Z_{m+k:m+k} \right\}. \quad (5.8)$$

We can write the above equation also as

$$U_{m+k}^{(m)} = \mathbf{b}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}}^{\prime} \mathbf{Z}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}}, \quad (5.9)$$

where $\mathbf{Z}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}} = (Z_{1:m+k}, Z_{2:m+k}, \dots, Z_{m+k:m+k})'$ and

$$\mathbf{b}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}}^{\prime} = \left(\sum_{j=1}^m \binom{m+k-1}{m-j} \binom{0}{j-1} c_{j:m}, \sum_{j=1}^m \binom{m+k-2}{m-j} \binom{1}{j-1} c_{j:m}, \dots, \sum_{j=1}^m \binom{0}{m-j} \binom{m+k-1}{j-1} c_{j:m} \right), \quad (5.10)$$

where in both (5.8) and (5.10) we define $\binom{p}{q} = 0$ for $p < q$. From (5.9) we write

$$\text{Var}(U_{m+k}^{(m)}) = (\mathbf{b}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}}^{\prime} \mathbf{G}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}} \mathbf{b}_{\underset{\sim}{\mathbf{m}+\mathbf{k}}}) \sigma^2, k = 1, 2, \dots, m - 1, \quad (5.11)$$

where $\mathbf{G}_{\mathbf{m}+\mathbf{k}}$ is the variance-covariance matrix of the vector of order statistics of a random sample of size $m+k$ drawn from the half-triangular distribution with pdf (3.3). Now from (5.7) and (5.11) we write

$$\begin{aligned} \binom{m}{m-k} \binom{k}{k} \xi_{m-k}^{(m)} + \binom{m}{m-k+1} \binom{k}{k-1} \xi_{m-k+1}^{(m)} + \dots + \binom{m}{m-1} \binom{k}{1} \xi_{m-1}^{(m)} \\ = \binom{m+k}{m} (\mathbf{b}_{\mathbf{m}+\mathbf{k}}' \mathbf{G}_{\mathbf{m}+\mathbf{k}} \mathbf{b}_{\mathbf{m}+\mathbf{k}}) \sigma^2 - \xi_m^{(m)} \end{aligned} \quad (5.12)$$

for, $k = 1, 2, \dots, m-1$, where we define for non-negative integers p and q , $\binom{p}{q} = 0$ for $p < q$. The above system of equations can be written by the following matrix equation:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & \binom{m}{m-1} \binom{1}{1} \\ 0 & 0 & \dots & \binom{m}{m-2} \binom{2}{2} & \binom{m}{m-1} \binom{2}{1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \binom{m}{1} \binom{m-1}{m-1} & \binom{m}{2} \binom{m-1}{m-2} & \dots & \binom{m}{m-2} \binom{m-1}{2} & \binom{m}{m-1} \binom{m-1}{1} \end{bmatrix} \begin{bmatrix} \xi_1^{(m)} \\ \xi_2^{(m)} \\ \vdots \\ \xi_{m-1}^{(m)} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{m-1} \end{bmatrix}, \quad (5.13)$$

where $\omega_k = \binom{m+k}{m} (\mathbf{b}_{\mathbf{m}+\mathbf{k}}' \mathbf{G}_{\mathbf{m}+\mathbf{k}} \mathbf{b}_{\mathbf{m}+\mathbf{k}}) \sigma^2 - \xi_m^{(m)}$, $k = 1, 2, \dots, m-1$.

If we write H to denote the coefficient matrix of the left side of (5.13) and $\underline{\omega}$ to denote the vector in the right side of (5.13), then we have

$$(\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_{m-1}^{(m)})' = H^{-1} \underline{\omega}.$$

The values of $\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_{m-1}^{(m)}$ can be obtained from the above equation and together with them if use $\xi_m^{(m)}$ as given in (5.2) and (5.6), we obtain the variance of the U-statistic $U_n^{(m)}$.

Thus as a result of the theory developed in this paper, we write the U-statistic using the BLUE of σ based on AOS as a kernel of degree m as $U_{n,A}^{(m)}$ which we trace out from (5.4) and is given by

$$U_{n,A}^{(m)} = \frac{1}{\binom{n}{m}} \sum_{r=1}^n \left[\sum_{j=1}^m \binom{n-r}{m-j} \binom{r-1}{j-1} c_{j:m} \right] X_{(r:n)}, \quad (5.14)$$

where, $X_{(1:n)}, X_{(2:n)}, \dots, X_{(n:n)}$ are the AOS of a random sample of size n drawn from the triangular distribution with pdf (1.1). The variance of $U_{n,A}^{(m)}$ is given by

$$Var[U_{n,A}^{(m)}] = \frac{1}{\binom{n}{m}} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \xi_c^{(m)}, \quad (5.15)$$

where $\xi_c^{(m)}, c = 1, 2, \dots, m-1$ are solved from (5.13) and $\xi_m^{(m)}$ (the variance of the kernel) is given as in (5.2). It is to be noted that provided $c_{j,m}$ for $j = 1, 2, \dots, m$ are determined then for any sample size n however large may be, one can obtain explicitly an unbiased, strongly consistent and asymptotically normal type of estimator $U_{n,A}^{(m)}$ for σ and is as given in (5.14). Similarly, if the means, variances and covariances of order statistics arising from half-triangular distribution for all sample sizes from m to $2m-1$ are obtained, then one can determine the exact variance of the estimator $U_{n,A}^{(m)}$ for any sample size however large it may be.

Now, in order to find that the minimal sufficient property of AOS arising from the triangular distribution with pdf (1.1) has induced any improvement on the U-statistic $U_{n,A}^{(m)}$, we obtain the variance $Var(U_{n,O}^{(m)})$ of a similar U-statistic $U_{n,O}^{(m)}$ constructed from the BLUE based on order statistics as kernel of the same degree arising from (1.1). If $Var(U_{n,A}^{(m)}) < Var(U_{n,O}^{(m)})$, then as usual we claim that $U_{n,A}^{(m)}$ is a better estimate of σ than the estimate $U_{n,O}^{(m)}$. Then the efficiency of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ is denoted by $e^{(m)}$ and is defined as $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$.

To illustrate the theory discussed above to estimate the scale parameter σ in (1.1), we consider BLUE's based on AOS with initial sample sizes $m = 2, 3, 4$ and 5 drawn from the triangular distribution and take them as kernels of degrees $2, 3, 4$ and 5 respectively. Using the estimate given by (3.8) and its linear representation given in (3.10), kernels in terms of order statistics arising from half-triangular distribution are derived and are given by

$$h(Z_1, Z_2) = 0.3333Z_{1:2} + 2.000Z_{2:2},$$

$$h(Z_1, Z_2, Z_3) = 0.1818Z_{1:3} + 0.2727Z_{2:3} + 1.6364Z_{3:3},$$

Table 5: Variances of U-statistics $U_{n,A}^{(m)}, U_{n,O}^{(m)}$ from triangular distribution and relative efficiencies $e^{(m)} = \frac{Var(U_{n,O}^{(m)})}{Var(U_{n,A}^{(m)})}$ for different n and m.

Sample Size	Variances of U-statistics										Relative efficiencies $e^{(m)}$		
	$U_{n,A}^{(2)}$	$U_{n,O}^{(2)}$	$U_{n,A}^{(3)}$	$U_{n,O}^{(3)}$	$U_{n,A}^{(4)}$	$U_{n,O}^{(4)}$	$U_{n,A}^{(5)}$	$U_{n,O}^{(5)}$	$e^{(2)}$	$e^{(3)}$	$e^{(4)}$	$e^{(5)}$	
10	0.0408	0.047	0.0371	0.0463	0.0349	0.0451	0.0334	0.0438	1.1520	1.2479	1.2923	1.3114	
15	0.0271	0.0295	0.0245	0.0291	0.0229	0.0283	0.0218	0.0274	1.0886	1.1878	1.2358	1.2569	
20	0.0203	0.0215	0.0183	0.0212	0.0171	0.0206	0.0161	0.0199	1.0591	1.1585	1.2047	1.2360	
30	0.0135	0.0140	0.0122	0.0137	0.0113	0.0133	0.0107	0.0129	1.0370	1.1229	1.1769	1.2056	
40	0.0101	0.01022	0.0091	0.0101	0.0085	0.0098	0.0079	0.0095	1.0189	1.1099	1.1529	1.2025	
60	0.0067	0.0068	0.0061	0.0067	0.0057	0.0065	0.0053	0.0062	1.0149	1.0984	1.1404	1.1698	
80	0.00503	0.0051	0.0045	0.0049	0.00425	0.0048	0.0041	0.0046	1.0139	1.0889	1.1429	1.1795	
100	0.004	0.0041	0.0036	0.0039	0.0034	0.0038	0.0032	0.0037	1.0250	1.0833	1.1176	1.1563	

$$h(Z_1, Z_2, Z_3, Z_4) = 0.1200Z_{1:4} + 0.1600Z_{2:4} + 0.2400Z_{3:4} + 1.4400Z_{4:4},$$

$$h(Z_1, Z_2, Z_3, Z_4, Z_5) = 0.0876Z_{1:5} + 0.1095Z_{2:5} + 0.146Z_{3:5} + 0.219Z_{4:5} + 1.3139Z_{5:5}.$$

Then using the above coefficients of the order statistics arising from half-triangular distribution in (5.14), the respective U-statistics $U_{n,A}^{(2)}$, $U_{n,A}^{(3)}$, $U_{n,A}^{(4)}$ and $U_{n,A}^{(5)}$ are obtained. Further, the components of variances $\xi_c^{(m)}$ for $c = 1, 2, \dots, m - 1$, $m = 2, 3, 4$ and 5 are computed and those values are tabulated in Table 4.

Using those values of $\xi_c^{(m)}$ for each of $m = 2, 3, 4, 5$ $\text{Var}(U_{n,A}^{(m)})$ for $n = 10(5)20(10)40(20)100$ are determined and are given in Table 5. As described in Thomas and Sreekumar (2008), $\text{Var}(U_{n,O}^{(m)})$ and efficiency $e^{(m)}$ of $U_{n,A}^{(m)}$ relative to $U_{n,O}^{(m)}$ for $n = 10(5)20(10)40(20)100$ are also calculated and these values are also presented in Table 5.

From Table 5, it can be seen that all efficiencies of U-statistics $U_{n,A}^{(m)}$ generated from BLUE based on AOS as kernels relative to the U-statistics $U_{n,O}^{(m)}$ generated from BLUE based on classical order statistics as kernels are greater than unity, on estimating the scale parameter of triangular distribution.

Acknowledgements: I would like to express my sincere gratitude to my research supervisors, Professor P. Yageen Thomas and Dr. Manoj Chacko, Department of Statistics, University of Kerala for their constant support, encouragement and positive suggestions in making this work possible.

References

- Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. John Wiley & Sons, New York.
- Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics & Inference: Estimation Methods*. Academic Press, New York.
- Balakrishnan, N. and Nevzorov, V. B. (2004). *A primer on statistical distributions*. John Wiley & Sons, New York.

- David, H. A. and Nagaraja, H. N. (2003). *Order Statistics, 3rd Edition*. John Wiley & Sons, New York.
- Fairchild, K. W., Misra, L., and Shi, Y. (2016). Using triangular distribution for business and finance simulations in excel. *Journal of Financial Education*, 42(3-4):313–336.
- Garg, M., Choudhary, S., and Nadarajah, S. (2009). On the product of triangular random variables. *Applicationes Mathematicae*, 36:419–439.
- Glickman, T. S. and Xu, F. (2008). The distribution of the product of two triangular random variables. *Statistics & Probability Letters*, 78(16):2821–2826.
- Johnson, D. (1997). The triangular distribution as a proxy for the beta distribution in risk analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(3):387–398.
- Johnson, D. (2002). Triangular approximations for continuous random variables in risk analysis. *Journal of the Operational Research Society*, 53(4):457–467.
- Johnson, N. L. and Kotz, S. (1999). Non-smooth sailing or triangular distributions revisited after some 50 years. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(2):179–187.
- Kotz, S. and René Van Dorp, J. (2004). *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*. World Scientific, Singapore.
- Rene van Dorp, J. and Kotz, S. (2002). A novel extension of the triangular distribution and its parameter estimation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(1):63–79.
- Samuel, P. and Thomas, P. Y. (2003). Estimation of the parameters of triangular distribution by order statistics. *Calcutta Statistical Association Bulletin*, 54(1-2):45–56.

- Sreekumar, N. V. and Thomas, P. Y. (2007). Estimation of the parameters of log-gamma distribution using order statistics. *Metrika*, 66(1):115–127.
- Sreekumar, N. V. and Thomas, P. Y. (2008). Estimation of the parameters of type-I generalized logistic distribution using order statistics. *Communications in Statistics-Theory and Methods*, 37(10):1506–1524.
- Thomas, P. Y. (1990). Estimating location and scale parameters of a symmetric distribution by systematic statistics. *Journal of Indian Society of Agriculture Statistics*, XLII(2):250–256.
- Thomas, P. Y. and Anjana, V. (2021). Estimation of the scale parameter of a family of distributions using a newly derived minimal sufficient statistic. *Communications in Statistics-Theory and Methods*, doi:10.1080/03610926.2021.1884721.
- Thomas, P. Y. and Baiju, K. V. (2012). Estimation of the scale parameter of skew-normal distribution using U-statistics based on order statistics. *Calcutta Statistical Association Bulletin*, 64(1-2):1–20.
- Thomas, P. Y. and Baiju, K. V. (2015). Estimation of the scale parameter of log-logistic distribution using U-statistics based on order statistics. *The Indian Association for Productivity, Quality and Reliability*, 40:111–128.
- Thomas, P. Y. and Priya, R. S. (2015). On a less cumbersome method of estimation of parameters of type iii generalized logistic distribution by order statistics. *Statistica*, 75(3):291–312.
- Thomas, P. Y. and Priya, R. S. (2016). Symmetric beta-Cauchy distribution and estimation of parameters using order statistics. *Calcutta Statistical Association Bulletin*, 68(1-2):111–134.
- Thomas, P. Y. and Samuel, P. (1996). A note on recurrence relations for the product moments of order statistics. *Statistics & Probability Letters*, 29(3):245–249.

Thomas, P. Y. and Sreekumar, N. V. (2008). Estimation of location and scale parameters of a distribution by U-statistics based on best linear functions of order statistics. *Journal of Statistical Planning and Inference*, 138(7):2190–2200.

L2 INPUT OF HIGH SCHOOL ESL LEARNERS IN THE INDIAN MILIEU: AN ANALYSIS**Dr Deepa L. C.**

Associate Professor of English, Govt. Arts College, Thiruvananthapuram.

Dr Josna Rajan

Assistant Professor of English, Government College, Kariavattom, Thiruvananthapuram.

Abstract

Communicative proficiency in English is indispensable for success today, especially in a Second Language (L2) context as in India. With the world shrinking to a global village, instant communication is the watch word, and English is the global language that links the various parts of this interconnected world. However, it is often found that long years of learning English language do not result in a desirable level of communicative ability in our learners. In the L2 situation prevailing in India, exposure to English is mainly through the input received through natural and instructional interaction in meaningful situations. This study attempts to analyze the nature and extent of the L2 input received by the learners of English at the high school level, with a view to modify this variable in such a way that at least a threshold level of communicative proficiency is ensured to the students by the time they enter colleges for higher studies. The study is highly relevant since L2 input is a variable that significantly influences learners' level of motivation which directly impacts the degree of communicative proficiency achieved by them.

Keywords: L1 (First Language or Mother tongue), L2 (Second Language), input, ESL (English as a Second Language), communicative proficiency.

Introduction

It is generally observed that even after years of learning English as a Second Language, a majority of Indian students are lacking in English language competence making even basic day to day communication difficult or impossible for them. English being a global language and a national link language in India, communicative proficiency in English is inevitable for survival. There are many factors that influence the L2 proficiency of the learners, of which L2 input plays a defining role.

Need and significance of the study

A large section of students who enroll for undergraduate courses in the colleges of India are found to fall short of the desired level of communicative proficiency in English. Therefore, it would be worthwhile to make an investigation into the L2 input received by the high school students of English, with a view to augment and modify it in order to ensure an optimum level of ideal input to enable them to attain the required level of proficiency in English by the time they enter colleges.

This study is an attempt to investigate the nature and extent of the L2 input received by the learners of English at the high school level, considering the role it plays in building their motivation to learn English, and its impact on the L2 competence acquired by the learners.

L2 input in ESL learning

The degree of motivation of a Second Language learner is largely impacted by the nature and intensity of the L2 input learners receive, which also determines how well they learn – both in L1 acquisition and L2 learning. A number of studies have emphasized the point that the ideal input for acquiring an L2 is similar to the input received by the child acquiring L1: comprehensible, relevant to one's immediate interests, not too complex but not strictly graded either (Dornyei 1994, Krashen 1989, Mc Laughlin, 1987). Thus it has been testified that in an L2 learning situation, the learner's natural acquisition mechanisms can be facilitated only through an exposure to such input.

Input in an L2 is received mainly from natural interaction or interaction in natural and meaningful situations, and instructional or classroom interaction. The present study examines the learners' extent of exposure to English through both natural and instructional interaction.

Natural Interaction

It is mainly the learner's natural learning mechanisms, that are activated only through actual and natural communication in the L2, that enable their L1 acquisition. Therefore, ideally, the learner should have abundant exposure to natural situations where the L2 is used for actual communication. Moreover, the learners should actively interact with native speakers in one-to-one communication. It is often observed that the longer one lives in the L2 country, the more proficient one becomes in the L2 (Littlewood, 58). Unfortunately, in the English language learning conditions prevailing in Kerala, interaction with native English speakers is scarcely possible. However, learners from the cities are more open to input in English since a section of the city inhabitants comprise employees who are not native to Kerala, but reside in Kerala to for their job purposes. Therefore, students may often have to rely on English for interacting with such families in their capacity as friends, neighbours and acquaintances. This indicates that natural interaction in English, within and outside the classroom is possible only for a limited number of pupils.

Instructional Interaction

In the Indian L2 learning milieu, the primary source of comprehensible input is classroom interaction – which includes teacher - student and student - student interaction in the classroom. A sufficient level of comprehensible input can ensure a minimum degree of L2 comprehension in the learners and help them engage in natural conversation at least to a certain extent. Collaborative activities such as discussion of topics of interest, games and tasks in keeping with students' age, interests, needs and backgrounds can go a long way in providing interesting and comprehensible input in the second language.

Interaction between teacher and students involves exchange of information, and it facilitates L2 acquisition. The teacher modifies his or her output to make it more comprehensible to the pupil, and the pupil progressively tries to make his or her own output understood by others. This kind of negotiations, if successful, provides the pupil with comprehensible input and an opportunity to use the target language, both of which are required for L2 development (Krashen cited in McLaughlin 45-46). This underscores the significance of both comprehensible input and comprehensible output in L2 development.

However, it is a sad commentary on the L2 learning environment of our state that the teacher often occupies the centre stage of the classroom explaining and describing the lessons. A sizeable section of the students consists of those lacking in effective communication skills who are at a loss to understand the teacher's words and hesitate to involve themselves in active communication in the classroom.

In order to improve the situation, learners of English at the primary, secondary and tertiary level, have to be given a lot of practice in using the language they have learnt. Communicative tasks such as simulation, role-play, projects, presentation and even answering question in class give them a chance to practise speaking English fluently. Moreover, the teacher should become a facilitator of learning in the classroom while students are provided with ample opportunities for active communicative interaction and collaborative group work. Students, not the teacher should do most part of the talking in the classroom. The teacher must no longer be a figure of authority, but an empathetic observer of the communicative activities, establishing a good rapport with the students and building/boosting their confidence to participate actively in the activities to their maximum. These measures will encourage healthy group dynamics among the peers, and there is a sense of joint responsibility running through them, a sense of "sinking or swimming together". (Kohonen 34).

Methodology and Procedure

The informants consist of a group of students and teachers of two government high schools in Thiruvananthapuram. Three hundred students of the tenth standard classes constituted the student informants. The teacher informants include eleven high school teachers of English from the two schools.

The methodology employed for the study consisted of survey followed by data analysis. The tools adopted for data collection include separate questionnaires for students and teachers, and interview with teachers. The questionnaire for students contained 25 questions which elicited data regarding the extent of the students' use of English in various situations in life – in family, with friends and neighbours, in social life, for academic activities, and for entertainment. The responses are expected to indicate the quality and quantity of L2 input they receive through natural interaction and instructional interaction. The questionnaire for teachers included 10 questions intended to gather information regarding the extent of L2 input provided to the students in the classroom. This was reinforced by the data gathered from interviews with the teachers.

The collected data was analysed to find out the degree of the L2 input received by the learners. A descriptive analysis of the data was carried out to determine the nature and extent of their exposure in various situations of everyday life.

Findings and Conclusions

The findings and conclusions of the study with regard to the L2 input received by the students in the classroom and outside are summarized as follows:

L2 input received through natural interaction

Input received from natural interaction in familiar situations pertains to the use of English in personal, social and academic life. Data was collected on the various situations in which students interact in English or receive input in English. It is found that English is used least in the *family* for communication. Only 7% of the informants use English for speaking, that too rarely. However, 12.8% use it for writing in the family circle.

English is used more for writing to friends and neighbours than for speaking to them. Only 13.2% of students use English occasionally for speaking to their *friends* while 38% use it for writing to them. For interacting with *neighbours*, English is used by 5% for speaking; and 25.9% for writing, mainly for social media messages.

English is used in *social life* by 8.2 % of students. The needs include ordering food at restaurants, talking to strangers and foreigners, talking to doctors, enquiries at the airport and the railway station, while shopping, for booking complaints over telephone etc.

It is to be noted that that a large section – 79 % of students use English for *academic or official purposes*. All the four skills are used in academic situations, and the speaking skill is used most widely for academic purposes than for personal and social purposes. The situations which mostly require the use of English are seminars, academic discussions, making notes, writing applications, reading academic books etc. Only 3.8% use English for talking to teachers outside the classroom.

It is welcoming that 62% of students use English in various situations for the purpose of *entertainment*. 39 % are exposed to English through reading newspapers and magazines, novels, stories, poems, religious books, etc. Input is received through the media by 45%, through sports commentary, news and recreational programmes on radio and TV, discussions and quiz . English is used to a large extent in the *electronic media*. 65% of the informants use English for sms, email, blogs, posts on social media and so on.

It is obvious that in most of the situations in which students use English, there is little scope for dynamic interaction in the language. Except in academic contexts, very few students use English for active communication. Thus, there is an acute dearth of opportunities for communication in the SL in real life. In other words, for a large majority, natural interaction in English is practically nil.

L2 input received through instructional interaction

Information regarding L2 input received through instructional interaction was collected by means of student questionnaire, teacher questionnaire and interviews with teachers. Students who actively interact in English with their teachers in the class constitute only 32%. Only around 6% use it for speaking, while 35 % use it for writing to teachers. Meanwhile, only a mere 15% of the informants interact with their classmates in English – 5% for speaking and 35% for writing. Thus, instructional interaction or active communication is found to be very low among the students in the classroom.

Responses elicited from the teachers through questionnaire and interviews about the approaches and methods used for teaching reinforce this finding. 92% of the teachers find the present teaching-learning situation unsatisfactory. There is little scope for interaction with students due to crowded classrooms, time constraints, unfriendly syllabus, lack of ICT and heavy workload.

Educational Implications

1. Collaborative group work must be encouraged in the classroom for creating opportunities for students for both natural and instructional interaction providing meaningful L2 input. Beyond mere talking, negotiation of meaning in communication is to be ensured.
2. The teacher must establish a good rapport with students creating a favourable classroom atmosphere that fosters good peer relationship and boosts confidence of learners.
3. Rather than being a provider of knowledge, the teacher should be a facilitator of learning.

4. Thus, it can be ensured that all students enter colleges for their undergraduate courses with confidence and at least a threshold level of proficiency in English .

Works Cited

Cooper, Nic A., and Betty K. Garner. *Developing a Learning Classroom: Moving beyond Management through Relationships, Relevance, and Rigor*. Corwin, 2012.

Dornyei, Zoltan. "Motivation and Motivating in the Foreign Language Classroom." *Modern Language Journal* 78.3 (1994): 273-84.

Kohonen, Viljo. "Experiential Language Learning: Second Language Learning as Cooperative Learner Education." *Collaborative Language Learning and Teaching*. Ed. David Nunan. Cambridge: Cambridge UP, 1992, 15 to 40.



Krashen, S. *Principles and Practices of Second Language Acquisition*. Oxford: Pergamon, 1982.

Littlewood, William T. *Foreign and Second Language Learning*. Cambridge: Cambridge UP, 1984.

McLaughlin, Barry. *Theories of Second Language Learning*. London: Edward Arnold, 1987, 52.

Review

Understanding Interaction Patterns within Deep-Sea Microbial Communities and Their Potential Applications

Muhammad Zohaib Nawaz¹, Raghul Subin Sasidharan², Huda Ahmed Alghamdi³ and Hongyue Dang^{1,*}

¹ State Key Laboratory of Marine Environmental Science, Fujian Key Laboratory of Marine Carbon Sequestration, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China; zohaib@xmu.edu.cn

² Department of Zoology, Government College Kariavattom, Thiruvananthapuram 695581, India; raghulzubin@gmail.com

³ Department of Biology, College of Sciences, King Khalid University, Abha 61413, Saudi Arabia; dr.huda.gh@gmail.com

* Correspondence: DangHY@xmu.edu.cn; Tel.: +86-592-2880151

Abstract: Environmental microbes living in communities engage in complex interspecies interactions that are challenging to decipher. Nevertheless, the interactions provide the basis for shaping community structure and functioning, which is crucial for ecosystem service. In addition, microbial interactions facilitate specific adaptation and ecological evolution processes particularly essential for microbial communities dwelling in resource-limiting habitats, such as the deep oceans. Recent technological and knowledge advancements provide an opportunity for the study of interactions within complex microbial communities, such as those inhabiting deep-sea waters and sediments. The microbial interaction studies provide insights into developing new strategies for biotechnical applications. For example, cooperative microbial interactions drive the degradation of complex organic matter such as chitins and celluloses. Such microbiologically-driven biogeochemical processes stimulate creative designs in many applied sciences. Understanding the interaction processes and mechanisms provides the basis for the development of synthetic communities and consequently the achievement of specific community functions. Microbial community engineering has many application potentials, including the production of novel antibiotics, biofuels, and other valuable chemicals and biomaterials. It can also be developed into biotechniques for waste processing and environmental contaminant bioremediation. This review summarizes our current understanding of the microbial interaction mechanisms and emerging techniques for inferring interactions in deep-sea microbial communities, aiding in future biotechnological and therapeutic applications.

Keywords: microbial interactions; microbial community; community function; correlation network analysis; deep sea



Citation: Nawaz, M.Z.; Subin Sasidharan, R.; Alghamdi, H.A.; Dang, H. Understanding Interaction Patterns within Deep-Sea Microbial Communities and Their Potential Applications. *Mar. Drugs* **2022**, *20*, 108. <https://doi.org/10.3390/md20020108>

Academic Editor: Orazio Tagliatalata-Scafati

Received: 6 November 2021

Accepted: 21 January 2022

Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Background

Considerable studies exploring microbial interactions at the community level have been done during the last decades [1–5]. Generally, the microorganisms living in a specific community may cooperate or compete for nutrients and other resources. They may exchange signal molecules and metabolites as well. Such interactions provide innate mechanisms in shaping the community structure, ecological function, and temporospatial dynamics of the microbiomes observed in various environments [6–8]. Intra- and interspecific interactions lay the foundation of the so-called “microbial community intelligence”, which can be explored for a variety of applications [9]. An interaction may exert a positive (win), negative (lose), or neutral (zero) impact on the individual microorganisms involved in the specific inter-species interaction [10–12]. Depending on the outcomes of the interaction between interacting species, the interaction can be classified into one of

several distinct situations [13], such as a win-win (mutualism) [14–17], win-lose (parasitism, predation) [18,19], win-zero (commensalism) [7,20], lose-lose (competition) [21], or a zero-lose (amensalism) relationship [10,22,23]. Although microbial interactions play important roles in driving the ocean's biogeochemical cycles [24] and the formation of coupled (or decoupled) community taxon-function dynamics in ecosystems [25], exploring the various types of interactions among the microorganisms in a complex community is not straightforward. Furthermore, interspecies interactions in microbial communities are not static, and evolution in interspecies interactions may occur over ecological timescales [26]. The interaction of evolution and ecology adds another layer of complexity to microbial interactions. Although it is a challenge to decode, the evolvability of microbial interactions contributes to the ecosystems' ecological memory and adaptive capacity, which may play critical roles in enabling the ecosystems to prepare for, and respond to, future perturbations such as the impacts of global change [27].

Recent time-series analyses have shown that some microbial communities may change in a resilient manner in response to environmental change (ecological resilience) [28], with the system attaining its original structure with time (engineering resilience) in response to species actions, including selection, dormancy, and speciation [29]. Species interactions may support bringing back the community's steady state in response to environmental perturbations (Figure 1) (e.g., changes in temperature, pH, oxygen concentration, redox potential, and nutrient supply) that trigger a change in the community structure [30]. Alternatively, some other microbial communities do not necessarily show resilience. Instead, they may tend to achieve an alternate stable state after a change in the environment [31]. Functional redundancy among distinct microbial species may provide a mechanism to maintain the community functionality with varied community compositions [32–34]. The effect of environmental perturbations on the microbial community structure has been illustrated in Figure 1. Microbiota in different environments may harbor varied taxonomic compositions. Nevertheless, they may host highly conserved community gene content and thus similar functional potentials [35]. Therefore, microbial communities can display one (mono-), two (bi-), or more (multi-) stable states under the same environmental conditions. The existence of a stable state(s) makes the microbial communities somehow fathomable [30,36]. It is possible to predict complex and dynamic interactions even within microbial communities in deep oceans [37].

The growing availability of microbial data in the marine environment reveals that the microbial interactions among species are more complex than previously thought. Emerging tools are being developed to infer such complex interactions. This review summarizes existing knowledge of the microbial interaction mechanisms and research tools for inferring relationships in deep-sea microbial communities, aiding future biotechnological and therapeutic applications.

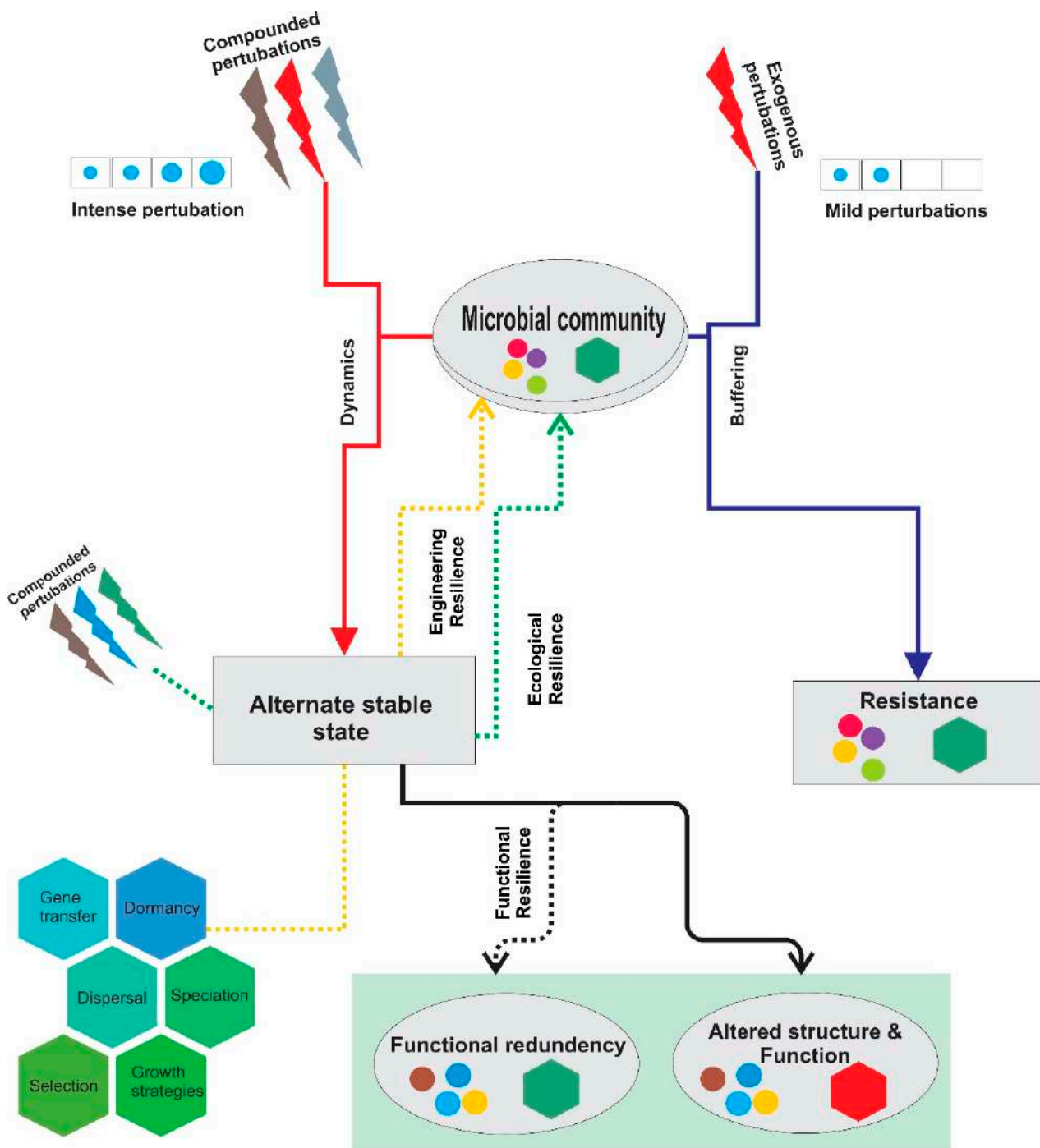


Figure 1. Effect of environmental disturbances on the microbial community composition.

2. The Complexity of Microbial Interaction in the Deep-Sea Environment

The deep-sea environments constitute vast and variable habitats for microorganisms, including viruses, archaea, bacteria, fungi, and protists. The deep-sea microorganisms usually form complex ecological interaction webs instead of dwelling in isolation. They are the key players in the deep-sea biogeochemical cycling of bio-essential elements, such as carbon, nitrogen, phosphorus, sulfur, and various trace metals [38,39]. A large number of microorganisms dwell in energy-deficient deep ocean sediments, which are considered the largest ecosystem on Earth [40]. Moreover, deep-sea hydrothermal vent chimneys characterized by steep physicochemical gradients harbor unique microbial communities that are particularly enriched with chemolithoautotrophic bacteria and archaea [41]. Sim-

ilarly, the deep-sea cold seeps also harbor one of the most productive ecosystems in the ocean, supporting complex microbial interactions centered on the coupling of anaerobic methane oxidation and sulfate reduction [42–44]. Due to the great demand for nitrogenous nutrients by the cold seep chemosynthetic ecosystems [45], nitrogen fixation by anaerobic methane-oxidizing archaea provides a critical mechanism to cope with the in-situ nitrogen deficiency [46–48]. Microbial interactions form the primary force driving the coupled cycling of carbon, nitrogen, sulfur, and other bio-essential elements in both hydrothermal vent and methane seep environments.

The seafloor deep biosphere is another extreme environment of the ocean. Due to the lack of sunlight and the extremely scarce supplies of organic matter from the surface ocean, the growth and eco-physiological activities of microorganisms living in the deep biosphere are highly limited by the meager availability of energy and organic substrates [49–51]. Under such resource-limited conditions, interspecies interactions such as metabolite cross-feeding and biosynthetic complementation may play a critical role for the in-situ microbial communities to fully exploit the available energy and growth substrates. Microorganisms carry out biochemically catalyzed redox reactions for energy transduction in the deep biosphere, where metabolically usable electron donors include methane, hydrogen, reduced iron, reduced manganese, reduced sulfur, ammonia, and ammonium. The electron acceptors in the deep biosphere include oxygen, oxidized nitrogen compounds such as nitrate and nitrite, manganese and iron oxides, oxidized sulfur compounds such as sulfate and sulfite, and oxidized carbon compounds such as carbon dioxide [52,53]. Different electron donors and acceptors are available in distinct habitats of the deep biosphere, forming the primary driving force to shape the community diversity, ecological function, and biogeography of the microorganisms inhabiting therein [51,54]. For example, in the oxic layer of the deep-sea sediments, aerobic microorganisms such as ammonia-oxidizing archaea and bacteria and nitrite-oxidizing bacteria may be the major chemolithoautotrophs contributing to inorganic carbon fixation [55]. In contrast, chemolithoautotrophic anaerobes contribute to the in-situ dark carbon fixation in the deep anoxic layers of sediments. The seafloor deep biosphere also represents other extreme conditions such as extreme temperature and high pressure. Despite improved knowledge of the microbial existence and diversity in the deep biosphere, mechanisms regarding habitat adaptation, metabolic activities, and interspecific interactions of the in-situ microbial communities remain largely elusive [56–58].

Studies have revealed through various molecular techniques not only the astonishing diversity but also the temporospatial dynamics of microbial abundance in deep-sea environments [59,60]. The microbial communities have usually been studied through phylogenetic analyses using taxonomic biomarkers such as 16S and 18S rRNA gene sequences [61,62]. Although these methods have made substantial contributions to the advancement of microbial ecology, they have certain limitations, including lineage missing caused by PCR primer mismatches and the inability of using single marker gene-based data to decode metabolic pathways and interactions in a microbial community [63]. Fortunately, these limitations have been overcome recently using multiple-omic analyses [64–67]. The deep-sea environments contain a vast diversity of microbial species and physiological traits [68], providing an opportunity for understanding microbial interactions in such ecologically and climatically critical earth subsystems.

Marine sediments contain a massive reservoir of living microorganisms, most of which may attach to sediment particles and live in biofilms thereon [69,70]. This surface-associated lifestyle may prompt various interactions among the sediment-dwelling microorganisms [5]. Microbes residing in biofilms are metabolically and functionally integrated microbial communities, displaying a high degree of organization and functioning as a unit with shared metabolites and signaling compounds [71]. Biofilms also facilitate gene expression regulation and horizontal gene transfer among community members [5]. Collective behavior of the microbial community is established by microbial interactions, such as those via the quorum sensing (QS)-based cell-to-cell communication mechanisms that allow the

interacting microbes to share information, materials, and functions [72]. QS communication is a response to microbial density that relies on the exchange of extracellular signaling molecules called autoinducers. QS enables microbial communities to behave like multicellular organisms, displaying mutual benefit, altruism, selfishness, or other social traits [73]. Microbes also possess other cell-to-cell communication mechanisms, such as the vesicle-mediated signal molecule transportation system [74] and the intercellular nanotubes-based microbial communication network [75]. Intra- and inter-species communications enable microbial communities to coordinate various ecophysiological processes, such as symbiosis, virulence, antibiotic production, and biofilm formation [76]. Engaging in social activities also enhances the survival of participating microbes in highly complex or adverse environments [71,77].

The interspecific interactions in sediment microbial communities have seldom been studied, particularly at the microscales that may provide the most relevant biological and ecological information about the in-situ microbial processes and mechanisms. The sediment microorganisms are essentially the engine driving the cycling of carbon and nutrients in the marine benthic ecosystems. They are also the cell factory carrying out biodegradation, biotransformation, and bioremediation of various contaminants and pollutants entering the ocean [78,79]. Marine sediments provide specific physicochemical and nutrient gradients, facilitating complex microbial communities and interspecific interactions. Seawater particles may provide similarly complex microenvironments, facilitating microbial interactions in the otherwise seemingly homogeneous bulk seawater environments [5]. The microbial interactions associated with seawater particulate organic matter play an important role in mediating the carbon sequestration efficiency of the biological carbon pump, a central mechanism played by the marine ecosystems for regulating the atmospheric CO₂ concentration and thus the climate [80]. Intensified microbial interactions associated with seawater particles and marine sediments may help the in-situ microbial communities evolve novel metabolic pathways and chemical compounds, presenting the potentials in biomedical, biotechnological, and industrial applications [81]. In highly challenging habitats such as the deep ocean environments, microorganisms existing as interacting community members may foster the ability to perform complex metabolic tasks via communicational cooperation and division of labor [82]. These ecological principles can be applied to the design and implementation of synthetic microbial communities for specific biomolecules and functions, revolutionizing the application of microbes and their interactions for therapeutic and biotechnological purposes [82,83].

Seawater sinking particles, such as macroscopic “marine snow” aggregates, act as a vector for transporting surface ocean-derived organic matter to the deep waters and seafloor [84]. Thus, they provide substantial organic substrates and nutrients to the deep ocean ecosystems. In addition, marine particles provide unique and partially isolated microenvironments by creating micro-scale suboxic or even anoxic conditions in the otherwise oxygenated bulk seawater of the ocean [5,85]. Diverse and interacting aerobic and anaerobic microbes colonize different niches of marine particles, taking advantage of the various organic and inorganic chemicals as growth substrates and electron donors and acceptors for energy metabolism (Figure 2). The productivity of the deep oceans may be highly influenced by the interaction of the particle-associated microbes, which form complex networks mainly characterized by substrate-level interdependencies. The heterotrophs seen in the aerobic and suboxic microzones of the marine particles (Figure 2) perform degradation of organic biopolymers, simultaneously consuming oxygen to maintain the redox gradient in the particles [5,55]. The abundances of particle-associated heterotrophs are usually orders of magnitude higher than those living in the surrounding seawater [86]. Microzonal oxygen-deficient conditions in the core of the marine snow particles facilitate anaerobic processes such as microbial denitrification, sulfate reduction, and methane production. The occurrence of ammonia and nitrite oxidizers in the oxic and suboxic microzones coupled with anaerobic ammonium oxidizers and denitrifiers in the anoxic microzones of the marine particles suggests an effective mechanism for the loss of fixed nitrogen in the ocean [87].

Similarly, coupled cycling of carbon, nitrogen, and sulfur was suggested in particular oxycline microzones of marine snow particles [85], where electron donors and acceptors may be actively recycled between distinct oxidation states (Figure 2). This network of biogeochemical interdependencies suggests the pivotal role of particle-associated microbes in the ecosystem metabolism of the deep ocean.

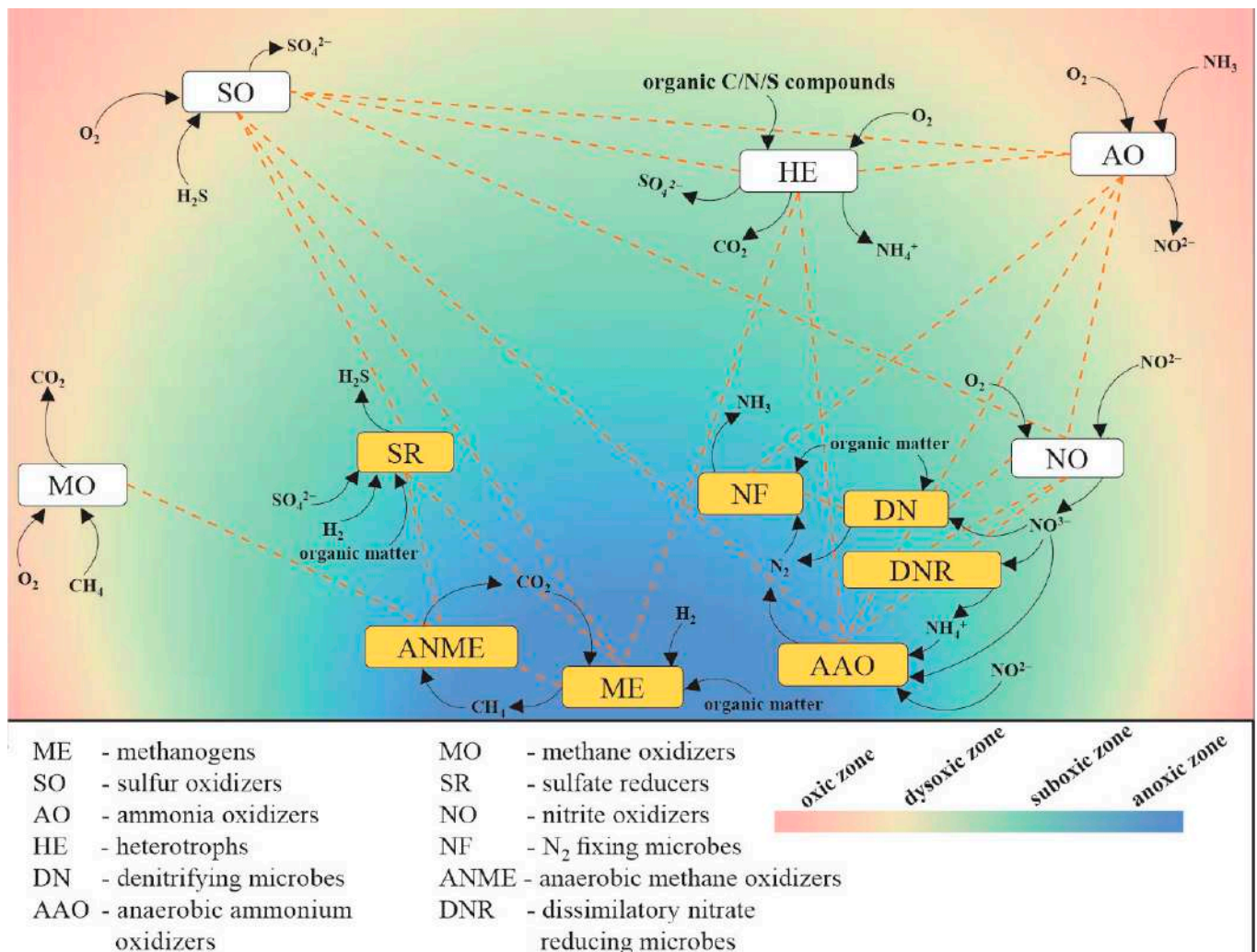


Figure 2. Interaction among C, N, and S cycling microbes associated with seawater particle aggregates in the deep-sea oxygenated waters.

3. Advancements in Molecular Techniques for Exploring Species Interaction

Species interactions in deep-sea microbial communities can be inferred either by analyzing their taxonomic data from different sampling sites at a given time or by analyzing time-series samples from the same sampling site. Moreover, an inferred interspecific relationship can be validated by investigating samples collected further from more habitats or from longer timescales [88,89]. As the availability of required samples from different environmental sites is usually not feasible, the time-series approach is more commonly opted for inferring species interactions. In this regard, ocean time-series study sites, such as the Bermuda Atlantic Time-series Study (BATS) site in the Sargasso Sea [90], the Hawaii ocean time-series (HOT) program Station ALOHA [91], and the South-East Asian Time-series Study (SEATS) site in the South China Sea [92], may prove valuable in providing ecologically meaningful materials and information for deep-sea microbial interaction analyses. The availability of time-series analysis tools has made it possible to develop predictive models and construct time-varying networks [93].

Experimental approaches for discovering interactions between species are mainly based on the concept of Gause's co-culture experiments [94], in which species interactions are tested by developing an artificial community in a controlled environment [95,96]. Although classical co-culture experiments can answer many ecological questions, they can only help reveal interactions among a limited number of species. Considerable progress has been made recently to extend the applications of co-culture experiments to include complex communities. However, there are still challenges, including identifying certain microbial community members, particularly those microbes that are difficult or even impossible to obtain pure strains.

Advancements in molecular technologies have made it possible to study complex microbial communities. For example, the development of PhyloChip and GeoChip for high-throughput co-culturing experiments may help reveal species interactions, even in deep-sea microbial communities [97]. Isotope labeling and probing experiments are helpful in deciphering the flow of metabolites and metabolic connections in microbial consortia [98]. Fluorescence in-situ hybridization (FISH)-based methods are helpful to explore the interaction pattern of co-aggregated species [98,99]. In combination with the microautoradiography (MAR) technique, the FISH technique holds the potential for observing the incorporation of radioisotopes in interacting microbial cells [100]. Similarly, the combination of FISH with Raman microscopy or high-resolution nanometer-scale secondary-ion mass spectrometry (NanoSIMS) can reveal microbial interactions with the use of stable isotope-labeled substrates [101–103].

The Deep-Sea Drilling Project made the first effort to collect and study microbial samples from various depths of marine sediments. Ever since these pioneering studies, growing molecular technologies have been used to study species interactions in deep-sea ecosystems. The first application of the combined FISH and secondary-ion mass spectrometry (SIMS) technique for exploring microbial interactions was reported by Orphan et al. [104], who studied methane-rich deep-sea sediments to determine the role of archaeal and bacterial cells in the anaerobic oxidation of methane. This study revealed the physical association of anaerobic methanotrophic archaea (ANME) with sulfate-reducing bacteria (SRB), suggesting that the interacting ANME and SRB consortia are responsible for the observed methanotrophy in anoxic cold seep sediments. Nowadays, next-generation "omics" approaches, including metagenomics, metatranscriptomics, metaproteomics, and metabolomics, have been being developed. These techniques hold great potential for the decoding of species interactions in deep-sea microbial communities [99,105].

4. Approaches for Exploring Species Interactions

4.1. Inferring Microbial Interactions through Co-Occurrence Pattern Analyses

Different approaches for exploring interspecific interactions in microbial communities have been developed, including co-occurrence pattern analyses, community metabolic pathway inferring, and eco-energetic modeling [13,106–108]. Co-occurring species may have similar ecological characteristics or may associate with each other because of physiological interdependencies or fine-scale niche differentiation. Thus, co-occurrence patterns may provide important insights into the temporospatial and functional distribution of microbes and the environmental complexity within an ecosystem. Microbial species with similar ecological traits can hardly co-exist in an environment when their common resources become limited. Competitive exclusion under resource-limiting conditions may eliminate some of the competing species that depend on the same limiting substrate for growth or survival [109]. However, species co-existence may stem from many distinct mechanisms, confounding the ecological explanation of an observed co-occurrence pattern. For example, Leinweber et al. recently proposed a "cheating effect" mechanism for fostering the co-existence of competitive species in a microbial community [110]. Under resource-limiting conditions, intraspecific competition caused by bacterial cheaters may alleviate interspecific competition, thus fostering species co-existence in a microbial community [110].

Co-occurrence patterns may be sensitive to different spatial scales or ecosystems. Williams et al. compared the co-occurrence data taken from different ecosystems, finding that only a few co-occurring pairs of microbial species showed consistency across different ecosystems. Most of the co-occurrence relationships detected in individual ecosystems were inconsistent across different ecosystems [111]. These results highlight the instability of using co-occurrence pattern analyses in inferring microbial interactions. Interactions among microbes are rooted in metabolic connections, providing a mechanistic and thus more reliable approach for detecting interacting microbes.

4.2. Inferring Microbial Interactions through Community Metabolic Pathway Analyses

The deep-sea ecosystem is the largest and most challenging ecosystem on Earth [112,113]. Most parts of this system are resource-limited. Recycling and reusing growth substrates and energy materials facilitate the maintenance and functioning of the deep-sea ecosystem, where the microorganisms may heavily rely on metabolite sharing to complement each other's biosynthetic requirements. Understanding metabolic exchanges and other forms of metabolic cooperation holds the key to decoding the microbial interactions therein. Community metabolic pathway analyses are challenging, particularly for the remote deep-sea microbial communities. Integrated utilization of diverse technologies, from metagenomics to isotope tracing, should facilitate the detection of metabolic pathways and networks in the deep-sea microbial communities [114].

Meta-omics techniques are powerful tools for identifying species and metabolic potentials in a microbial community and inferring interactions among different community members. Data generated by these techniques can be used to formulate and test new hypotheses about microbial metabolic interactions (Figure 3). A microbial community as a whole can be treated as a super- or mega-organism, whose metabolic pathways and networks are fathomable with the use of the -omics approaches [115,116]. For example, cooperation or competition among microbes may be inferred taxonomically from time-series analysis of a specific community [117]. Meta-omics analyses with emphases on the microbial functional traits may be applied subsequently to corroborate further or contradict the inference made by taxonomy-based analyses [13,118,119]. Metagenomics data usually reveal the metabolic potentials of a microbial community, while metatranscriptomics and metaproteomics data may reveal more about the metabolic activities of the studied microbial community [120,121]. The molecular sequence-based -omics techniques can be further combined with isotope tracing techniques to explore the flow of metabolites in a microbial community [122]. Advanced microscopy imaging techniques, such as cryogenic transmission electron microscopy, have been developed to visualize physical interactions among microbial cells in complex microbial communities [123]. Imaging mass spectrometry has also been developed to visualize interspecies metabolic exchange between interacting microbial cells [124,125]. These cutting-edge techniques will undoubtedly produce more exciting discoveries in deep-sea microbial interaction studies.

Advancements in molecular techniques, particularly the—omics ones, have improved our understanding of microbial ecology, providing the essential data and concepts for mathematical modeling to predict metabolic interactions of the various microbial communities [126–128]. Sequenced genomes facilitate the construction of the metabolic network in single organisms by providing the necessary information of the metabolic enzymes and the biochemical reactions they catalyze [129]. Community metabolic networks can be constructed using the microbial community's metagenomic data [130]. From a constructed metabolic network, it is convenient to identify the metabolites that are not synthesized by a microorganism but may be obtained from other community members [131,132]. Potential substrate competitions among different microorganisms can also be identified in a community metabolic network. Therefore, community metabolic networks provide critical information about the species-specific resource requirements and metabolic cooperation and competition in a given microbial community [130,133], providing valuable insights into the processes and mechanisms of microbial interactions (Figure 4).

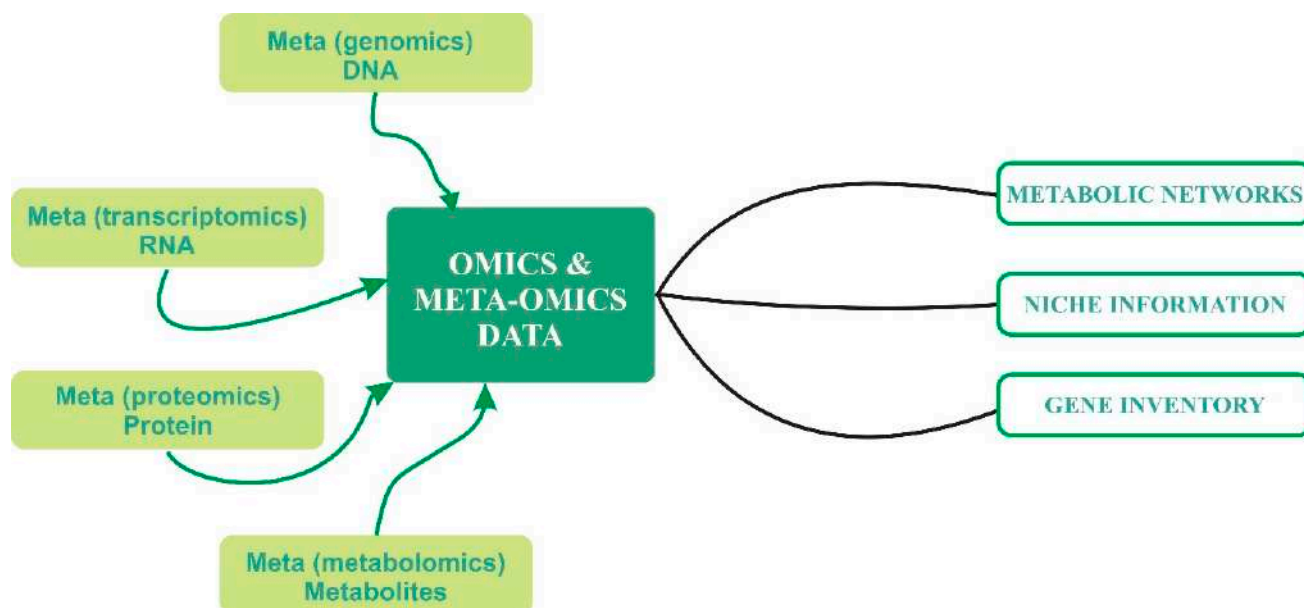


Figure 3. Meta-omics approaches for studying microbial communities and their functions.

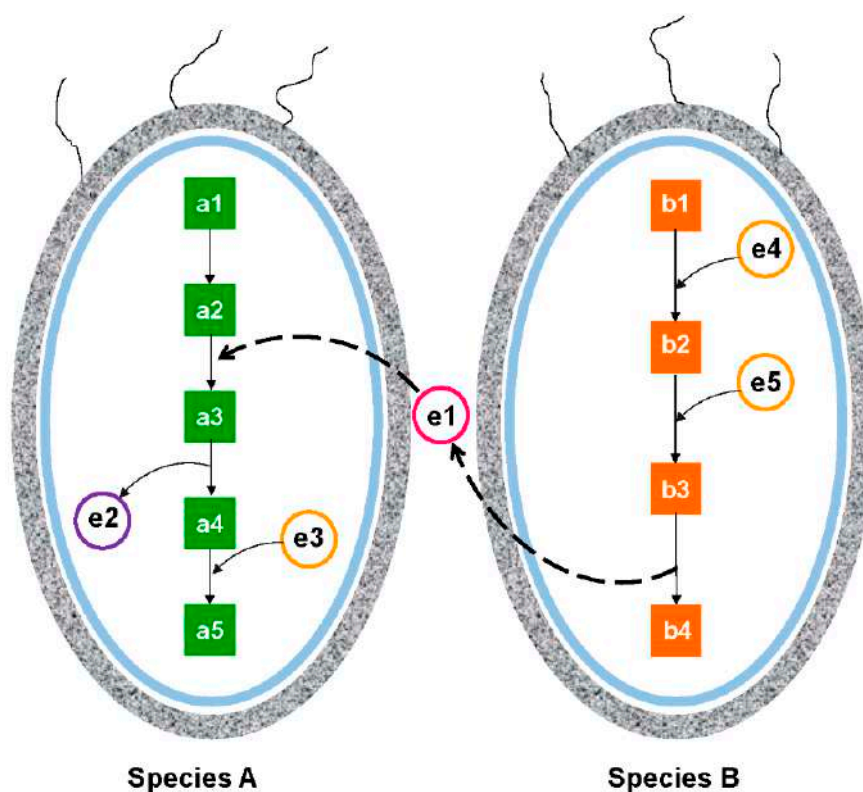


Figure 4. Two species in a microbial community interact with each other by sharing metabolites to fuel each other's metabolic pathways. Metabolic pathway of species A needs metabolite “e1” that is not synthesized by its own metabolic machinery and needs to be taken up from species B through the environment. Therefore, species A must coexist with species B that synthesizes this metabolite. Components a1, a2, a3, a4, and a5 represent the genes in the metabolic pathway of species A, whereas b1, b2, b3, and b4 are genes in the metabolic pathway of species B. Components e1, e2, e3, e4, and e5 are different intermediate metabolites produced or needed by the two species.

Metabolic networks also provide information about the community's metabolic environment [134]. The habitat conditions and corresponding microbial adaptation mechanisms may be inferred from metabolic network details [131,132,135,136]. Valuable information about the habitat attributes and microbial adaptation strategies may be used to design culture media to isolate interested microorganisms.

4.3. *Inferring Microbial Interactions through Community Eco-Energetic Modeling*

Microbes dwelling the deep-sea environments conserve energy via various redox reactions. The coupling of oxidation and reduction reactions between different microbes facilitates microbial interactions along various energetic substrate gradients [55,137,138]. For example, marine sediments are rich in metal oxides and other minerals that serve as electron donors or electron acceptors in many microbiologically-catalyzed redox reactions [139,140]. Although the microbial cell envelope forms a permeability barrier to minerals, many microbes have evolved extracellular electron transfer mechanisms for using minerals to exchange electrons [140–143]. Marine mineral-mediated redox reactions thus can directly or indirectly facilitate microbial interspecific electron transfer and thus microbial interactions [144]. Redox coupling without mineral involvement also prevails in natural microbial communities. For example, aggregates or other forms of consortia formed by different microbes are common in marine environments. They facilitate otherwise difficult microbial metabolism, such as anaerobic oxidation of methane (AOM) [145,146]. Facilitated interspecific electron transfer between methane-oxidizing archaea and sulfate-reducing bacteria has been proposed as the primary mechanism for AOM in the aggregates [147,148].

Different energetic substrates (i.e., electron donors and electron acceptors) exist in distinct marine environments. The specific pairing of the available electron donors and their counterpart electron acceptors dictates the eco-energetic processes and the functional groups of microbes and the interaction pattern that can occur in a given environment [55,149,150]. Here, we propose that understanding the electron transfer interactions between different microbial species would help reveal some major interaction mechanisms and environmental adaptation strategies of deep-sea microorganisms. Combining the -omics techniques and eco-energetic modeling may prove fruitful to achieve this aim.

4.4. *Synthetic Microbial Communities in Biotechnological and Therapeutic Applications*

Interspecific interactions enable microbes to survive in highly challenging environments. Metabolic cooperation helps the deep-sea microbes maximize utilizing the available resources, including recalcitrant organic substrates [151], such as cellulose, lignin, chitin, lipids, and hydrocarbons. For example, syntrophic interactions between Lokiarchaeota and nitrite- or sulfite-reducing bacteria may enable these archaea to anaerobically degrade aliphatic and aromatic hydrocarbons in marine subsurface sediments [152]. On the contrary, acetogenic Bathyarchaeota may help fuel the marine subsurface ecosystem by providing organic substrates for heterotrophy and acetoclastic methanogenesis [40].

The capability of deep-sea microbes for collaborative degradation of complex and recalcitrant organic compounds may have many applications, such as biofuel production from cellulose, lignin, and other polysaccharide substances. The organic-decomposing communities and their enzymes are also useful for developing bioremediation techniques for coping with various environmental contaminants [153]. In addition, metagenomics-based approaches also reveal a plethora of enzymes that may have other potential applications, such as in the energy, biomedical, industrial, and biotechnological fields [154]. Enzymes from extremophiles (e.g., thermophiles and psychrophiles) deserve special attention because extremozymes' higher stability and catalytic efficiency may have many therapeutic and biotechnological potentials [155,156].

Opportunistic pathogens usually lead to polymicrobial infections. An improved understanding of their colonization mechanisms and cell-to-cell communications is necessary for developing effective therapeutic strategies for diseases caused by multispecies infections [157]. The microbial QS-based cell-to-cell communication system is a promising target for interfering microbial interactions to prevent pathogen colonization, biofilm formation, and polymicrobial infections [82,158]. Some marine sediment bacteria secrete AHL-lactonase enzymes that can disrupt QS-mediated microbial interactions [159,160]. The QS-disrupting enzymes and the microbes producing them hold promising potential for tackling biofilm formation and infections by bacterial pathogens. A practical and potentially effective strategy is to engineer microbial symbioses involving AHL-lactonase-producing microbes to control bacterial infections [161]. Engineered microbial symbionts may also be applied to invade and kill cancerous cells [162]. Furthermore, autoinducer antagonists such as QS inhibitors may be developed and employed as effective antimicrobial drugs [163]. The deep-sea microorganisms may harbor many novel bio-chemicals and mechanisms that can interfere with microbial interactions, potentially applicable in therapeutics [164].

5. Future Directions

Species interactions are an important force shaping community structure and function. Species interactions can also strongly impact how organisms and their communities respond to global warming, ocean acidification, ocean deoxygenation, and other environmental stressors [165]. Inferring species interactions of the deep-sea microbial communities is still a challenging task. However, technological integrations, particularly those involving combined -omics, isotope tracing, and modeling approaches, make high-quality research possible in this field. New concepts and innovative theories may be further developed to achieve an ecosystems biology vision of the complex marine microbial communities [166,167]. Understanding the microbial interactions provides a gateway to the design of synthetic microbial communities to obtain novel or unique community functions.

Deep-sea microbial interaction research has been entering a golden age of rapid development since the turn of the new millennium. It is reasonable to predict that new scientific discoveries and theoretical advancements will thrive, which may help advance the applications of interspecific interaction-based biotechniques in many fields, such as therapeutics and bioremediation. For example, the use of synthetic bioremediation communities at the industrial scale has several advantages over the traditional genetic engineering approaches (Figure 5), such as greater adaptability in extreme conditions and lesser adverse effect on the ecology of the in situ microbial communities. Exploring the interactions in the deep-sea microbial communities opens new horizons for the advancement of microbial ecology and the applications of microbial theories for therapeutic and biotechnological purposes.

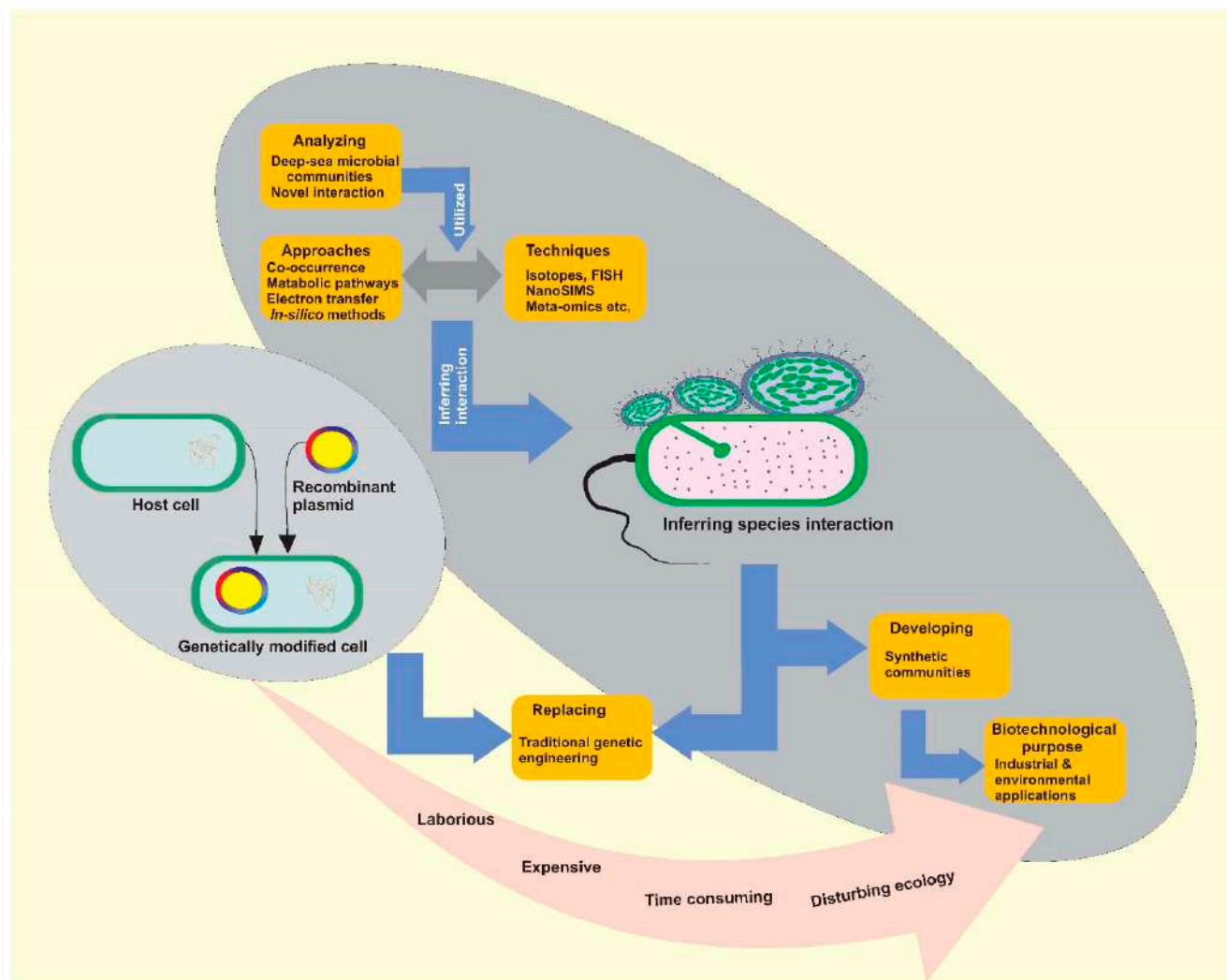


Figure 5. Advantages of synthetic communities' applications over traditional genetic engineering at industrial scale.

Funding: This study was supported by the National Key Research and Development Program of China (grant No. 2020YFA0608302), the National Natural Science Foundation of China (Nos. 42076111, 42141003, 42188102, 41861144018, and 41676122), and the China Ocean Mineral Resources R&D Association (No. DY135-E2-1-04).

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Wintermute, E.H.; Silver, P.A. Dynamics in the mixed microbial concourse. *Genes Dev.* **2010**, *24*, 2603–2614. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Xavier, J.B. Social interaction in synthetic and natural microbial communities. *Mol. Syst. Biol.* **2011**, *7*, 483. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Tanouchi, Y.; Smith, R.P.; You, L. Engineering microbial systems to explore ecological and evolutionary dynamics. *Curr. Opin. Biotechnol.* **2012**, *23*, 791–797. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Mitri, S.; Foster, K.R. The genotypic view of social interactions in microbial communities. *Annu. Rev. Genet.* **2013**, *47*, 247–273. [\[CrossRef\]](#)
5. Dang, H.; Lovell, C.R. Microbial surface colonization and biofilm development in marine environments. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 91–138. [\[CrossRef\]](#)
6. Follows, M.J.; Dutkiewicz, S.; Grant, S.; Chisholm, S.W. Emergent biogeography of microbial communities in a model ocean. *Science* **2007**, *315*, 1843–1846. [\[CrossRef\]](#)
7. Freilich, S.; Kreimer, A.; Meilijson, I.; Gophna, U.; Sharan, R.; Ruppin, E. The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.* **2010**, *38*, 3857–3868. [\[CrossRef\]](#)

8. Steele, J.A.; Countway, P.D.; Xia, L.; Vigil, P.D.; Beman, J.M.; Kim, D.Y.; Chow, C.E.; Sachdeva, R.; Jones, A.C.; Schwalbach, M.S.; et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* **2011**, *5*, 1414–1425. [[CrossRef](#)]
9. Tikariha, H.; Purohit, H.J. Unfolding microbial community intelligence in aerobic and anaerobic biodegradation processes using metagenomics. *Arch. Microbiol.* **2020**, *202*, 1269–1274. [[CrossRef](#)]
10. Konopka, A. What is microbial community ecology? *ISME J.* **2009**, *3*, 1223–1230. [[CrossRef](#)]
11. Ghouil, M.; Mitri, S. The ecology and evolution of microbial competition. *Trends Microbiol.* **2016**, *24*, 833–845. [[CrossRef](#)] [[PubMed](#)]
12. Pande, S.; Kost, C. Bacterial unculturability and the formation of intercellular metabolic networks. *Trends Microbiol.* **2017**, *25*, 349–361. [[CrossRef](#)] [[PubMed](#)]
13. Perez-Garcia, O.; Lear, G.; Singhal, N. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front. Microbiol.* **2016**, *7*, 673. [[CrossRef](#)] [[PubMed](#)]
14. Shou, W.; Ram, S.; Vilar, J.M. Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1877–1882. [[CrossRef](#)] [[PubMed](#)]
15. Hillesland, K.L.; Stahl, D.A. Rapid evolution of stability and productivity at the origin of a microbial mutualism. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2124–2129. [[CrossRef](#)]
16. Summers, Z.M.; Fogarty, H.E.; Leang, C.; Franks, A.E.; Malvankar, N.S.; Lovley, D.R. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. *Science* **2010**, *330*, 1413–1415. [[CrossRef](#)]
17. Kerner, A.; Park, J.; Williams, A.; Lin, X.N. A programmable *Escherichia coli* consortium via tunable symbiosis. *PLoS ONE* **2012**, *7*, e34032. [[CrossRef](#)]
18. Kerr, B.; Riley, M.A.; Feldman, M.W.; Bohannan, B.J. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **2002**, *418*, 171–174. [[CrossRef](#)]
19. Balagadde, F.K.; Song, H.; Ozaki, J.; Collins, C.H.; Barnet, M.; Arnold, F.H.; Quake, S.R.; You, L. A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.* **2008**, *4*, 187. [[CrossRef](#)]
20. Xu, A.; Dolfig, J.; Curtis, T.P.; Montague, G.; Martin, E. Maintenance affects the stability of a two-tiered microbial ‘food chain’? *J. Theor. Biol.* **2011**, *276*, 35–41. [[CrossRef](#)]
21. Foster, K.R.; Bell, T. Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* **2012**, *22*, 1845–1850. [[CrossRef](#)] [[PubMed](#)]
22. Diamond, J.M.; Cody, M.L. *Ecology and Evolution of Communities*; Belknap Press of Harvard University Press: Cambridge, MA, USA, 1975.
23. Lidicker, W.Z. A clarification of interactions in ecological systems. *BioScience* **1979**, *29*, 475–477. [[CrossRef](#)]
24. Corno, G.; Salka, I.; Pohlmann, K.; Hall, A.R.; Grossart, H.P. Interspecific interactions drive chitin and cellulose degradation by aquatic microorganisms. *Aquat. Microb. Ecol.* **2015**, *76*, 27–37. [[CrossRef](#)]
25. Lima-Mendez, G.; Faust, K.; Henry, N.; Decelle, J.; Colin, S.; Carcillo, F.; Chaffron, S.; Ignacio-Espinosa, J.C.; Roux, S.; Vincent, F. Determinants of community structure in the global plankton interactome. *Science* **2015**, *348*. [[CrossRef](#)] [[PubMed](#)]
26. Gorter, F.A.; Manhart, M.; Ackermann, M. Understanding the evolution of interspecies interactions in microbial communities. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2020**, *375*, 20190256. [[CrossRef](#)]
27. Angeler, D.G.; Fried-Petersen, H.B.; Allen, C.R.; Garmestani, A.; Twidwell, D.; Chuang, W.-C.; Donovan, V.M.; Eason, T.; Roberts, C.P.; Sundstrom, S.M. Adaptive capacity in ecosystems. In *Advances in Ecological Research*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 60, pp. 1–24.
28. Shade, A.; Read, J.S.; Youngblut, N.D.; Fierer, N.; Knight, R.; Kratz, T.K.; Lottig, N.R.; Roden, E.E.; Stanley, E.H.; Stombaugh, J.; et al. Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME J.* **2012**, *6*, 2153–2167. [[CrossRef](#)]
29. Philippot, L.; Griffiths, B.S.; Langenheder, S. Microbial community resilience across ecosystems and multiple disturbances. *Microbiol. Mol. Biol. Rev.* **2021**, *85*, e00026–20. [[CrossRef](#)]
30. Costello, E.K.; Stagaman, K.; Dethlefsen, L.; Bohannan, B.J.; Relman, D.A. The application of ecological theory toward an understanding of the human microbiome. *Science* **2012**, *336*, 1255–1262. [[CrossRef](#)]
31. Allison, S.D.; Martiny, J.B. Colloquium paper: Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (Suppl. 1), 11512–11519. [[CrossRef](#)]
32. Bissett, A.; Burke, C.; Cook, P.L.; Bowman, J.P. Bacterial community shifts in organically perturbed sediments. *Environ. Microbiol.* **2007**, *9*, 46–60. [[CrossRef](#)]
33. Yannarell, A.C.; Steppe, T.F.; Paerl, H.W. Disturbance and recovery of microbial community structure and function following Hurricane Frances. *Environ. Microbiol.* **2007**, *9*, 576–583. [[CrossRef](#)] [[PubMed](#)]
34. Miki, T.; Yokokawa, T.; Matsui, K. Biodiversity and multifunctionality in a microbial community: A novel theoretical approach to quantify functional redundancy. *Proc. Biol. Sci.* **2014**, *281*, 20132498. [[CrossRef](#)] [[PubMed](#)]
35. Louca, S.; Polz, M.F.; Mazel, F.; Albright, M.B.N.; Huber, J.A.; O’Connor, M.I.; Ackermann, M.; Hahn, A.S.; Srivastava, D.S.; Crowe, S.A.; et al. Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2018**, *2*, 936–943. [[CrossRef](#)] [[PubMed](#)]
36. Fuhrman, J.A.; Cram, J.A.; Needham, D.M. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **2015**, *13*, 133–146. [[CrossRef](#)]
37. Little, A.E.; Robinson, C.J.; Peterson, S.B.; Raffa, K.F.; Handelsman, J. Rules of engagement: Interspecies interactions that regulate microbial communities. *Annu. Rev. Microbiol.* **2008**, *62*, 375–401. [[CrossRef](#)]

38. Falkowski, P.G.; Fenchel, T.; Delong, E.F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **2008**, *320*, 1034–1039. [[CrossRef](#)]
39. Zehr, J.P.; Kudela, R.M. Nitrogen cycle of the open ocean: From genes to ecosystems. *Ann. Rev. Mar. Sci.* **2011**, *3*, 197–225. [[CrossRef](#)]
40. He, Y.; Li, M.; Perumal, V.; Feng, X.; Fang, J.; Xie, J.; Sievert, S.M.; Wang, F. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat. Microbiol.* **2016**, *1*, 16035. [[CrossRef](#)]
41. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 805–814. [[CrossRef](#)]
42. Jorgensen, B.B.; Boetius, A. Feast and famine—microbial life in the deep-sea bed. *Nat. Rev. Microbiol.* **2007**, *5*, 770–781. [[CrossRef](#)]
43. Levin, L.A.; Sibuet, M. Understanding continental margin biodiversity: A new imperative. *Ann. Rev. Mar. Sci.* **2012**, *4*, 79–112. [[CrossRef](#)] [[PubMed](#)]
44. Bienhold, C.; Pop Ristova, P.; Wenzhofer, F.; Dittmar, T.; Boetius, A. How deep-sea wood falls sustain chemosynthetic life. *PLoS ONE* **2013**, *8*, e53590. [[CrossRef](#)] [[PubMed](#)]
45. Fulweiler, R.W. Microbiology. Fantastic fixers. *Science* **2009**, *326*, 377–378. [[CrossRef](#)] [[PubMed](#)]
46. Dang, H.; Luan, X.; Zhao, J.; Li, J. Diverse and novel *nifH* and *nifH*-like gene sequences in the deep-sea methane seep sediments of the Okhotsk Sea. *Appl. Environ. Microbiol.* **2009**, *75*, 2238–2245. [[CrossRef](#)]
47. Dekas, A.E.; Poretsky, R.S.; Orphan, V.J. Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* **2009**, *326*, 422–426. [[CrossRef](#)]
48. Miyazaki, J.; Higa, R.; Toki, T.; Ashi, J.; Tsunogai, U.; Nunoura, T.; Imachi, H.; Takai, K. Molecular characterization of potential nitrogen fixation by anaerobic methane-oxidizing archaea in the methane seep sediments at the number 8 Kumano Knoll in the Kumano Basin, offshore of Japan. *Appl. Environ. Microbiol.* **2009**, *75*, 7153–7162. [[CrossRef](#)]
49. Hoehler, T.M.; Jorgensen, B.B. Microbial life under extreme energy limitation. *Nat. Rev. Microbiol.* **2013**, *11*, 83–94. [[CrossRef](#)]
50. Orsi, W.D. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat. Rev. Microbiol.* **2018**, *16*, 671–683. [[CrossRef](#)]
51. Bradley, J.; Arndt, S.; Amend, J.; Burwicz, E.; Dale, A.W.; Egger, M.; LaRowe, D.E. Widespread energy limitation to life in global subseafloor sediments. *Sci. Adv.* **2020**, *6*, eaba0697. [[CrossRef](#)]
52. McCollom, T.M. Geochemical constraints on primary productivity in submarine hydrothermal vent plumes. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2000**, *47*, 85–101. [[CrossRef](#)]
53. Bach, W.; Edwards, K.J. Iron and sulfide oxidation within the basaltic ocean crust: Implications for chemolithoautotrophic microbial biomass production. *Geochim. Cosmochim. Acta* **2003**, *67*, 3871–3887. [[CrossRef](#)]
54. Graw, M.F.; D'Angelo, G.; Borchers, M.; Thurber, A.R.; Johnson, J.E.; Zhang, C.; Liu, H.; Colwell, F.S. Energy gradients structure microbial communities across sediment horizons in deep marine sediments of the South China sea. *Front. Microbiol.* **2018**, *9*, 729. [[CrossRef](#)] [[PubMed](#)]
55. Dang, H.; Chen, C.T.A. Ecological energetic perspectives on responses of nitrogen-transforming chemolithoautotrophic microbiota to changes in the marine environment. *Front. Microbiol.* **2017**, *8*, 1246. [[CrossRef](#)] [[PubMed](#)]
56. Harcombe, W. Novel cooperation experimentally evolved between species. *Evolution* **2010**, *64*, 2166–2172. [[CrossRef](#)] [[PubMed](#)]
57. Wintermute, E.H.; Silver, P.A. Emergent cooperation in microbial metabolism. *Mol. Syst. Biol.* **2010**, *6*, 407. [[CrossRef](#)] [[PubMed](#)]
58. Hom, E.F.; Murray, A.W. Plant-fungal ecology. Niche engineering demonstrates a latent capacity for fungal-algal mutualism. *Science* **2014**, *345*, 94–98. [[CrossRef](#)]
59. Boeuf, D.; Edwards, B.R.; Eppley, J.M.; Hu, S.K.; Poff, K.E.; Romano, A.E.; Caron, D.A.; Karl, D.M.; DeLong, E.F. Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11824–11832. [[CrossRef](#)]
60. Baker, B.J.; Appler, K.E.; Gong, X. New microbial biodiversity in marine sediments. *Ann. Rev. Mar. Sci.* **2021**, *13*, 161–175. [[CrossRef](#)]
61. Sogin, M.L.; Morrison, H.G.; Huber, J.A.; Mark Welch, D.; Huse, S.M.; Neal, P.R.; Arrieta, J.M.; Herndl, G.J. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12115–12120. [[CrossRef](#)]
62. Bartram, A.K.; Lynch, M.D.; Stearns, J.C.; Moreno-Hagelsieb, G.; Neufeld, J.D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* **2011**, *77*, 3846–3852. [[CrossRef](#)]
63. Zhou, J.; Deng, Y.; He, Z.; Wu, L.; Van Nostrand, J.D. Applying GeoChip analysis to disparate microbial communities. *Microbe Mag.* **2010**, *5*, 60–65. [[CrossRef](#)]
64. Rondon, M.R.; August, P.R.; Bettermann, A.D.; Brady, S.F.; Grossman, T.H.; Liles, M.R.; Loiacono, K.A.; Lynch, B.A.; MacNeil, I.A.; Minor, C.; et al. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **2000**, *66*, 2541–2547. [[CrossRef](#)] [[PubMed](#)]
65. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–685. [[CrossRef](#)] [[PubMed](#)]

66. Tyson, G.W.; Chapman, J.; Hugenholtz, P.; Allen, E.E.; Ram, R.J.; Richardson, P.M.; Solovyev, V.V.; Rubin, E.M.; Rokhsar, D.S.; Banfield, J.F. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **2004**, *428*, 37–43. [[CrossRef](#)] [[PubMed](#)]
67. Tringe, S.G.; von Mering, C.; Kobayashi, A.; Salamov, A.A.; Chen, K.; Chang, H.W.; Podar, M.; Short, J.M.; Mathur, E.J.; Detter, J.C.; et al. Comparative metagenomics of microbial communities. *Science* **2005**, *308*, 554–557. [[CrossRef](#)] [[PubMed](#)]
68. Wang, F.; Lu, S.; Orcutt, B.N.; Xie, W.; Chen, Y.; Xiao, X.; Edwards, K.J. Discovering the roles of subsurface microorganisms: Progress and future of deep biosphere investigation. *Sci. Bull.* **2012**, *58*, 456–467. [[CrossRef](#)]
69. Meyer-Reil, L.-A. Microbial life in sedimentary biofilms—The challenge to microbial ecologists. *Mar. Ecol. Prog. Ser.* **1994**, *112*, 303–311. [[CrossRef](#)]
70. Flemming, H.C.; Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* **2019**, *17*, 247–260. [[CrossRef](#)]
71. Berlanga, M.; Guerrero, R. Living together in biofilms: The microbial cell factory and its biotechnological implications. *Microb. Cell Fact* **2016**, *15*, 165. [[CrossRef](#)]
72. Williams, P. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **2007**, *153*, 3923–3938. [[CrossRef](#)]
73. Diggle, S.P. Microbial communication and virulence: Lessons from evolutionary theory. *Microbiology* **2010**, *156*, 3503–3512. [[CrossRef](#)] [[PubMed](#)]
74. Mashburn-Warren, L.M.; Whiteley, M. Special delivery: Vesicle trafficking in prokaryotes. *Mol. Microbiol.* **2006**, *61*, 839–846. [[CrossRef](#)] [[PubMed](#)]
75. Dubey, G.P.; Ben-Yehuda, S. Intercellular nanotubes mediate bacterial communication. *Cell* **2011**, *144*, 590–600. [[CrossRef](#)]
76. Schauder, S.; Bassler, B.L. The languages of bacteria. *Genes Dev.* **2001**, *15*, 1468–1480. [[CrossRef](#)] [[PubMed](#)]
77. Hooshangi, S.; Bentley, W.E. From unicellular properties to multicellular behavior: Bacteria quorum sensing circuitry and applications. *Curr. Opin. Biotechnol.* **2008**, *19*, 550–555. [[CrossRef](#)] [[PubMed](#)]
78. Bedard, D.L. A case study for microbial biodegradation: Anaerobic bacterial reductive dechlorination of polychlorinated biphenyls—from sediment to defined medium. *Annu. Rev. Microbiol.* **2008**, *62*, 253–270. [[CrossRef](#)] [[PubMed](#)]
79. Zhang, D.; Li, X.; Zhang, C.; Xiao, Z.; Li, Y.; Liang, Y.; Dang, H. Electrostimulated bio-dechlorination of a PCB mixture (Aroclor 1260) in a marine-originated dechlorinating culture. *Environ. Pollut.* **2021**, *291*, 118157. [[CrossRef](#)]
80. Cavan, E.L.; Laurenceau-Cornec, E.C.; Bressac, M.; Boyd, P.W. Exploring the ecology of the mesopelagic biological pump. *Prog. Oceanogr.* **2019**, *176*, 102125. [[CrossRef](#)]
81. Ambrosino, L.; Tangherlini, M.; Colantuono, C.; Esposito, A.; Sangiovanni, M.; Miralto, M.; Sansone, C.; Chiusano, M.L. Bioinformatics for marine products: An overview of resources, bottlenecks, and perspectives. *Mar. Drugs* **2019**, *17*, 576. [[CrossRef](#)]
82. Brenner, K.; You, L.; Arnold, F.H. Engineering microbial consortia: A new frontier in synthetic biology. *Trends Biotechnol.* **2008**, *26*, 483–489. [[CrossRef](#)]
83. Ben Said, S.; Or, D. Synthetic microbial ecology: Engineering habitats for modular consortia. *Front. Microbiol.* **2017**, *8*, 1125. [[CrossRef](#)] [[PubMed](#)]
84. Alldredge, A.L.; Silver, M.W. Characteristics, dynamics and significance of marine snow. *Prog. Oceanogr.* **1988**, *20*, 41–82. [[CrossRef](#)]
85. Wright, J.J.; Konwar, K.M.; Hallam, S.J. Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **2012**, *10*, 381–394. [[CrossRef](#)] [[PubMed](#)]
86. Simon, M.; Grossart, H.P.; Schweitzer, B.; Ploug, H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat. Microb. Ecol.* **2002**, *28*, 175–211. [[CrossRef](#)]
87. Bristow, L.A. Anoxia in the snow. *Nat. Geosci.* **2018**, *11*, 226–227. [[CrossRef](#)]
88. Berry, D.; Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **2014**, *5*, 219. [[CrossRef](#)]
89. Zelezniak, A.; Andrejev, S.; Ponomarova, O.; Mende, D.R.; Bork, P.; Patil, K.R. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6449–6454. [[CrossRef](#)]
90. Vergin, K.L.; Jhirad, N.; Dodge, J.; Carlson, C.A.; Giovannoni, S.J. Marine bacterioplankton consortia follow deterministic, non-neutral community assembly rules. *Aquat. Microb. Ecol.* **2017**, *79*, 165–175. [[CrossRef](#)]
91. Karl, D.M. Microbial oceanography: Paradigms, processes and promise. *Nat. Rev. Microbiol.* **2007**, *5*, 759–769. [[CrossRef](#)]
92. Wong, G.T.F.; Ku, T.-L.; Mulholland, M.; Tseng, C.-M.; Wang, D.-P. The SouthEast Asian Time-series Study (SEATS) and the biogeochemistry of the South China Sea—An overview. *Deep Sea Res. Part II Top. Stud.* **2007**, *54*, 1434–1447. [[CrossRef](#)]
93. Faust, K.; Lahti, L.; Gonze, D.; de Vos, W.M.; Raes, J. Metagenomics meets time series analysis: Unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **2015**, *25*, 56–66. [[CrossRef](#)] [[PubMed](#)]
94. Gause, G.F. Experimental analysis of Vito Volterra’s mathematical theory of the struggle for existence. *Science* **1934**, *79*, 16–17. [[CrossRef](#)]
95. Grosskopf, T.; Soyer, O.S. Synthetic microbial communities. *Curr. Opin. Microbiol.* **2014**, *18*, 72–77. [[CrossRef](#)] [[PubMed](#)]
96. Song, H.; Ding, M.Z.; Jia, X.Q.; Ma, Q.; Yuan, Y.J. Synthetic microbial consortia: From systematic analysis to construction and applications. *Chem. Soc. Rev.* **2014**, *43*, 6954–6981. [[CrossRef](#)] [[PubMed](#)]
97. Faust, K.; Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **2012**, *10*, 538–550. [[CrossRef](#)] [[PubMed](#)]

98. Fowler, S.J.; Gutierrez-Zamora, M.L.; Manefield, M.; Gieg, L.M. Identification of toluene degraders in a methanogenic enrichment culture. *FEMS Microbiol. Ecol.* **2014**, *89*, 625–636. [[CrossRef](#)] [[PubMed](#)]
99. Behrens, S.; Losekann, T.; Pett-Ridge, J.; Weber, P.K.; Ng, W.O.; Stevenson, B.S.; Hutcheon, I.D.; Relman, D.A.; Spormann, A.M. Linking microbial phylogeny to metabolic activity at the single-cell level by using enhanced element labeling-catalyzed reporter deposition fluorescence in situ hybridization (EL-FISH) and NanoSIMS. *Appl. Environ. Microbiol.* **2008**, *74*, 3143–3150. [[CrossRef](#)]
100. Nielsen, P.H.; Andreassen, K.; Lee, N.; Wagner, M. Use of microautoradiography and fluorescent hybridization for characterization of microbial activity in activated sludge. *Water Sci. Technol.* **1999**, *39*, 1–9. [[CrossRef](#)]
101. Huang, W.E.; Stoecker, K.; Griffiths, R.; Newbold, L.; Daims, H.; Whiteley, A.S.; Wagner, M. Raman-FISH: Combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. *Environ. Microbiol.* **2007**, *9*, 1878–1889. [[CrossRef](#)]
102. Guerquin-Kern, J.-L.; Wu, T.-D.; Quintana, C.; Croisy, A. Progress in analytical imaging of the cell by dynamic secondary ion mass spectrometry (SIMS microscopy). *Biochim. Biophys. Acta* **2005**, *1724*, 228–238. [[CrossRef](#)]
103. Wagner, M. Single-cell ecophysiology of microbes as revealed by Raman microspectroscopy or secondary ion mass spectrometry imaging. *Annu. Rev. Microbiol.* **2009**, *63*, 411–429. [[CrossRef](#)] [[PubMed](#)]
104. Orphan, V.J.; House, C.H.; Hinrichs, K.U.; McKeegan, K.D.; DeLong, E.F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7663–7668. [[CrossRef](#)] [[PubMed](#)]
105. Hallam, S.J.; Mincer, T.J.; Schleper, C.; Preston, C.M.; Roberts, K.; Richardson, P.M.; DeLong, E.F. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol.* **2006**, *4*, e95. [[CrossRef](#)] [[PubMed](#)]
106. Raes, J.; Bork, P. Molecular eco-systems biology: Towards an understanding of community function. *Nat. Rev. Microbiol.* **2008**, *6*, 693–699. [[CrossRef](#)] [[PubMed](#)]
107. Chaffron, S.; Rehrauer, H.; Pernthaler, J.; von Mering, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **2010**, *20*, 947–959. [[CrossRef](#)] [[PubMed](#)]
108. Eiler, A.; Heinrich, F.; Bertilsson, S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* **2012**, *6*, 330–342. [[CrossRef](#)]
109. Kalenitchenko, D.; Fagervold, S.K.; Pruski, A.M.; Vétion, G.; Yucel, M.; Le Bris, N.; Galand, P.E. Temporal and spatial constraints on community assembly during microbial colonization of wood in seawater. *ISME J.* **2015**, *9*, 2657–2670. [[CrossRef](#)]
110. Leinweber, A.; Fredrik Inglis, R.; Kummerli, R. Cheating fosters species co-existence in well-mixed bacterial communities. *ISME J.* **2017**, *11*, 1179–1188. [[CrossRef](#)]
111. Williams, R.J.; Howe, A.; Hofmockel, K.S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* **2014**, *5*, 358. [[CrossRef](#)]
112. Parkes, R.J.; Cragg, B.A.; Bale, S.J.; Getliff, J.M.; Goodman, K.; Rochelle, P.A.; Fry, J.C.; Weightman, A.J.; Harvey, S.M. Deep bacterial biosphere in Pacific Ocean sediments. *Nature* **1994**, *371*, 410–413. [[CrossRef](#)]
113. Whitman, W.B.; Coleman, D.C.; Wiebe, W.J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6578–6583. [[CrossRef](#)] [[PubMed](#)]
114. Ponomarova, O.; Patil, K.R. Metabolic interactions in microbial communities: Untangling the Gordian knot. *Curr. Opin. Microbiol.* **2015**, *27*, 37–44. [[CrossRef](#)] [[PubMed](#)]
115. Lykidis, A.; Chen, C.L.; Tringe, S.G.; McHardy, A.C.; Copeland, A.; Kyrpides, N.C.; Hugenholtz, P.; Macarie, H.; Olmos, A.; Monroy, O.; et al. Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *ISME J.* **2011**, *5*, 122–130. [[CrossRef](#)] [[PubMed](#)]
116. Shafquat, A.; Joice, R.; Simmons, S.L.; Huttenhower, C. Functional and phylogenetic assembly of microbial communities in the human microbiome. *Trends Microbiol.* **2014**, *22*, 261–266. [[CrossRef](#)] [[PubMed](#)]
117. Ju, F.; Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* **2015**, *9*, 683–695. [[CrossRef](#)] [[PubMed](#)]
118. Lawson, C.E.; Wu, S.; Bhattacharjee, A.S.; Hamilton, J.J.; McMahon, K.D.; Goel, R.; Noguera, D.R. Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Commun.* **2017**, *8*, 15416. [[CrossRef](#)] [[PubMed](#)]
119. Ofaim, S.; Ofek-Lalzar, M.; Sela, N.; Jinag, J.; Kashi, Y.; Minz, D.; Freilich, S. Analysis of microbial functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. *Front. Microbiol.* **2017**, *8*, 1606. [[CrossRef](#)]
120. Embree, M.; Nagarajan, H.; Movahedi, N.; Chitsaz, H.; Zengler, K. Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. *ISME J.* **2014**, *8*, 757–767. [[CrossRef](#)]
121. Hawley, A.K.; Brewer, H.M.; Norbeck, A.D.; Pasa-Tolic, L.; Hallam, S.J. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11395–11400. [[CrossRef](#)]
122. Verastegui, Y.; Cheng, J.; Engel, K.; Kolczynski, D.; Mortimer, S.; Lavigne, J.; Montalibet, J.; Romantsov, T.; Hall, M.; McConkey, B.J.; et al. Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. *mBio* **2014**, *5*, e01157-14. [[CrossRef](#)]
123. Comolli, L.R.; Banfield, J.F. Inter-species interconnections in acid mine drainage microbial communities. *Front. Microbiol.* **2014**, *5*, 367. [[CrossRef](#)] [[PubMed](#)]
124. Shih, C.-J.; Chen, P.-Y.; Liaw, C.-C.; Lai, Y.-M.; Yang, Y.-L. Bringing microbial interactions to light using imaging mass spectrometry. *Nat. Prod. Rep.* **2014**, *31*, 739–755. [[CrossRef](#)] [[PubMed](#)]

125. Chen, Q.; Wang, L.; Qi, Y.; Ma, C. Imaging mass spectrometry of interspecies metabolic exchange revealed the allelopathic interaction between *Microcystis aeruginosa* and its antagonist. *Chemosphere* **2020**, *259*, 127430. [[CrossRef](#)] [[PubMed](#)]
126. Klitgord, N.; Segre, D. Environments that induce synthetic microbial ecosystems. *PLoS. Comput. Biol.* **2010**, *6*, e1001002. [[CrossRef](#)]
127. Mahadevan, R.; Henson, M.A. Genome-based modeling and design of metabolic interactions in microbial communities. *Comput. Struct. Biotechnol. J.* **2012**, *3*, e201210008. [[CrossRef](#)]
128. Zengler, K.; Palsson, B.O. A road map for the development of community systems (CoSy) biology. *Nat. Rev. Microbiol.* **2012**, *10*, 366–3672. [[CrossRef](#)]
129. Feist, A.M.; Herrgard, M.J.; Thiele, I.; Reed, J.L.; Palsson, B.O. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **2009**, *7*, 129–143. [[CrossRef](#)]
130. Cardona, C.; Weisenhorn, P.; Henry, C.; Gilbert, J.A. Network-based metabolic analysis and microbial community modeling. *Curr. Opin. Microbiol.* **2016**, *31*, 124–131. [[CrossRef](#)]
131. Borenstein, E.; Kupiec, M.; Feldman, M.W.; Ruppin, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14482–14487. [[CrossRef](#)]
132. Handorf, T.; Christian, N.; Ebenhoh, O.; Kahn, D. An environmental perspective on metabolism. *J. Theor. Biol.* **2008**, *252*, 530–537. [[CrossRef](#)]
133. Borenstein, E.; Feldman, M.W. Topological signatures of species interactions in metabolic networks. *J. Comput. Biol.* **2009**, *16*, 191–200. [[CrossRef](#)]
134. Roling, W.F.; Ferrer, M.; Golyshin, P.N. Systems approaches to microbial communities and their functioning. *Curr. Opin. Biotechnol.* **2010**, *21*, 532–538. [[CrossRef](#)] [[PubMed](#)]
135. Morine, M.J.; Gu, H.; Myers, R.A.; Bielawski, J.P. Trade-offs between efficiency and robustness in bacterial metabolic networks are associated with niche breadth. *J. Mol. Evol.* **2009**, *68*, 506–515. [[CrossRef](#)] [[PubMed](#)]
136. Freilich, S.; Kreimer, A.; Borenstein, E.; Yosef, N.; Sharan, R.; Gophna, U.; Ruppin, E. Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol.* **2009**, *10*, R61. [[CrossRef](#)] [[PubMed](#)]
137. Shi, L.; Dong, H.; Reguera, G.; Beyenal, H.; Lu, A.; Liu, J.; Yu, H.Q.; Fredrickson, J.K. Extracellular electron transfer mechanisms between microorganisms and minerals. *Nat. Rev. Microbiol.* **2016**, *14*, 651–662. [[CrossRef](#)]
138. Le Bris, N.; Yücel, M.; Das, A.; Sievert, S.M.; LokaBharathi, P.; Girguis, P.R. Hydrothermal energy transfer and organic carbon production at the deep seafloor. *Front. Mar. Sci.* **2019**, *5*, 531. [[CrossRef](#)]
139. Wu, W.F.; Wang, F.P.; Li, J.H.; Yang, X.W.; Xiao, X.; Pan, Y.X. Iron reduction and mineralization of deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3 at elevated hydrostatic pressures. *Geobiology* **2013**, *11*, 593–601.
140. Melton, E.D.; Swanner, E.D.; Behrens, S.; Schmidt, C.; Kappler, A. The interplay of microbially mediated and abiotic reactions in the biogeochemical Fe cycle. *Nat. Rev. Microbiol.* **2014**, *12*, 797–808. [[CrossRef](#)]
141. Liu, Y.; Wang, Z.; Liu, J.; Levar, C.; Edwards, M.J.; Babauta, J.T.; Kennedy, D.W.; Shi, Z.; Beyenal, H.; Bond, D.R.; et al. A trans-outer membrane porin-cytochrome protein complex for extracellular electron transfer by *Geobacter sulfurreducens* PCA. *Environ. Microbiol. Rep.* **2014**, *6*, 776–785. [[CrossRef](#)]
142. Bird, L.J.; Bonnefoy, V.; Newman, D.K. Bioenergetic challenges of microbial iron metabolisms. *Trends Microbiol.* **2011**, *19*, 330–340. [[CrossRef](#)]
143. Reguera, G.; McCarthy, K.D.; Mehta, T.; Nicoll, J.S.; Tuominen, M.T.; Lovley, D.R. Extracellular electron transfer via microbial nanowires. *Nature* **2005**, *435*, 1098–1101. [[CrossRef](#)] [[PubMed](#)]
144. Kato, S.; Hashimoto, K.; Watanabe, K. Microbial interspecies electron transfer via electric currents through conductive minerals. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10042–10046. [[CrossRef](#)] [[PubMed](#)]
145. Knittel, K.; Boetius, A. Anaerobic oxidation of methane: Progress with an unknown process. *Annu. Rev. Microbiol.* **2009**, *63*, 311–334. [[CrossRef](#)] [[PubMed](#)]
146. Thauer, R.K. Anaerobic oxidation of methane with sulfate: On the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO₂. *Curr. Opin. Microbiol.* **2011**, *14*, 292–299. [[CrossRef](#)] [[PubMed](#)]
147. McGlynn, S.E.; Chadwick, G.L.; Kempes, C.P.; Orphan, V.J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **2015**, *526*, 531–535. [[CrossRef](#)]
148. Wegener, G.; Krukenberg, V.; Riedel, D.; Tegetmeyer, H.E.; Boetius, A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **2015**, *526*, 587–590. [[CrossRef](#)]
149. Dodsworth, J.A.; McDonald, A.I.; Hedlund, B.P. Calculation of total free energy yield as an alternative approach for predicting the importance of potential chemolithotrophic reactions in geothermal springs. *FEMS Microbiol. Ecol.* **2012**, *81*, 446–454. [[CrossRef](#)]
150. Bradley, J.A.; Amend, J.P.; LaRowe, D.E. Bioenergetic controls on microbial ecophysiology in marine sediments. *Front. Microbiol.* **2018**, *9*, 180. [[CrossRef](#)]
151. Sebastian, M.; Forn, I.; Auladell, A.; Gomez-Letona, M.; Sala, M.M.; Gasol, J.M.; Marrase, C. Differential recruitment of opportunistic taxa leads to contrasting abilities in carbon processing by bathypelagic and surface microbial communities. *Environ. Microbiol.* **2021**, *23*, 190–206. [[CrossRef](#)]
152. Farag, I.F.; Biddle, J.F.; Zhao, R.; Martino, A.J.; House, C.H.; Leon-Zayas, R.I. Metabolic potentials of archaeal lineages resolved from metagenomes of deep Costa Rica sediments. *ISME J.* **2020**, *14*, 1345–1358. [[CrossRef](#)]
153. Dombrowski, N.; Seitz, K.W.; Teske, A.P.; Baker, B.J. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* **2017**, *5*, 106. [[CrossRef](#)] [[PubMed](#)]

154. Zhang, C.; Kim, S.K. Research and application of marine microbial enzymes: Status and prospects. *Mar. Drugs* **2010**, *8*, 1920–1934. [[CrossRef](#)] [[PubMed](#)]
155. Raddadi, N.; Cherif, A.; Daffonchio, D.; Neifar, M.; Fava, F. Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7907–7913. [[CrossRef](#)]
156. Sarmiento, F.; Peralta, R.; Blamey, J.M. Cold and hot extremozymes: Industrial relevance and current trends. *Front. Bioeng. Biotechnol.* **2015**, *3*, 148. [[CrossRef](#)] [[PubMed](#)]
157. Tshikantwa, T.S.; Ullah, M.W.; He, F.; Yang, G. Current trends and potential applications of microbial interactions for human welfare. *Front. Microbiol.* **2018**, *9*, 1156. [[CrossRef](#)] [[PubMed](#)]
158. Waters, C.M.; Bassler, B.L. Quorum sensing: Cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 319–346. [[CrossRef](#)]
159. Rehman, Z.U.; Leiknes, T. Quorum-quenching bacteria isolated from Red Sea sediments reduce biofilm formation by *Pseudomonas aeruginosa*. *Front. Microbiol.* **2018**, *9*, 1354. [[CrossRef](#)]
160. Packiavathy, I.; Kannappan, A.; Thiagarajan, S.; Srinivasan, R.; Jeyapragash, D.; Paul, J.B.J.; Velmurugan, P.; Ravi, A.V. AHL-lactonase producing *Psychrobacter sp.* from Palk Bay sediment mitigates quorum sensing-mediated virulence production in Gram negative bacterial pathogens. *Front. Microbiol.* **2021**, *12*, 634593. [[CrossRef](#)]
161. Hays, S.G.; Patrick, W.G.; Ziesack, M.; Oxman, N.; Silver, P.A. Better together: Engineering and application of microbial symbioses. *Curr. Opin. Biotechnol.* **2015**, *36*, 40–49. [[CrossRef](#)]
162. Anderson, J.C.; Clarke, E.J.; Arkin, A.P.; Voigt, C.A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* **2006**, *355*, 619–627. [[CrossRef](#)]
163. Chifiriuc, M.C.; Grumezescu, A.M.; Lazar, V. Quorum sensing inhibitors from the sea: Lessons from marine symbiotic relationships. *Curr. Org. Chem.* **2014**, *18*, 823–839. [[CrossRef](#)]
164. Marmann, A.; Aly, A.H.; Lin, W.; Wang, B.; Proksch, P. Co-cultivation—a powerful emerging tool for enhancing the chemical diversity of microorganisms. *Mar. Drugs* **2014**, *12*, 1043–1065. [[CrossRef](#)] [[PubMed](#)]
165. Gilman, S.E.; Urban, M.C.; Tewksbury, J.; Gilchrist, G.W.; Holt, R.D. A framework for community interactions under climate change. *Trends Ecol. Evol.* **2010**, *25*, 325–331. [[CrossRef](#)]
166. Pacheco, A.R.; Segre, D. A multidimensional perspective on microbial interactions. *FEMS Microbiol. Lett.* **2019**, *366*, fnz125. [[CrossRef](#)] [[PubMed](#)]
167. Sunagawa, S.; Acinas, S.G.; Bork, P.; Bowler, C.; Tara Oceans, C.; Eveillard, D.; Gorsky, G.; Guidi, L.; Iudicone, D.; Karsenti, E.; et al. Tara Oceans: Towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **2020**, *18*, 428–445. [[CrossRef](#)]



Appraisal of cytotoxicity and acrylamide mitigation potential of L-asparaginase SlpA from fish gut microbiome

Tina Kollannoor Johny¹ · Rinu Madhu Puthusseri¹ · Bindiya Ellathuparambil Saidumohamed¹ · Unnikrishnan Babukuttan Sheela² · Saipriya Parol Puthusseri¹ · Raghul Subin Sasidharan³ · Sarita Ganapathy Bhat¹

Received: 18 December 2021 / Revised: 27 April 2022 / Accepted: 3 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

L-asparaginase catalyzes the hydrolysis of L-asparagine to L-aspartic acid and ammonia. It has application in the treatment of acute lymphoblastic leukemia in children, as well as in other malignancies, in addition to its role as a food processing aid for the mitigation of acrylamide formation in the baking industry. Its use in cancer chemotherapy is limited due to problems such as its intrinsic glutaminase activity and associated side effects, leading to an increased interest in the search for novel L-asparaginases without L-glutaminase activity. This study reports the cloning and expression of an L-asparaginase contig obtained from whole metagenome shotgun sequencing of *Sardinella longiceps* gut microbiota. Purified recombinant glutaminase-free L-asparaginase SlpA was a 74 kDa homodimer, with maximal activity at pH 8 and 30 °C. K_m and V_{max} of SlpA were determined to be 3.008 mM and 0.014 mM/min, respectively. SlpA displayed cytotoxic activity against K-562 (chronic myeloid leukemia) and MCF-7 (breast cancer) cell lines with IC_{50} values of 0.3443 and 2.692 U/mL, respectively. SlpA did not show any cytotoxic activity against normal lymphocytes and was proved to be hemocompatible. Pre-treatment of biscuit and bread dough with different concentrations of SlpA resulted in a clear, dose-dependent reduction of acrylamide formation during baking.

Key points

- Cloned and expressed L-asparaginase (SlpA) from fish gut microbiota
- Purified SlpA displayed good cytotoxicity against K-562 and MCF-7 cell lines
- SlpA addition caused a significant reduction of acrylamide formation during baking

Keywords L-asparaginase · Anticancer · Leukemia · Acrylamide mitigation · Baking

✉ Sarita Ganapathy Bhat
saritagbhat@gmail.com

Tina Kollannoor Johny
tina.kj05@gmail.com

Rinu Madhu Puthusseri
rinumadhu@gmail.com

Bindiya Ellathuparambil Saidumohamed
bindiya79@yahoo.co.in

Unnikrishnan Babukuttan Sheela
ukbsbio@gmail.com

Saipriya Parol Puthusseri
pp.priya.p@gmail.com

Raghul Subin Sasidharan
raghulzubin@gmail.com

¹ Department of Biotechnology, Cochin University of Science and Technology, Cochin, Kerala 682022, India

² Department of Plant Biotechnology, College of Agriculture, Vellayani, Kerala 695522, India

³ Department of Zoology, Government College Kariavattom, University of Kerala, Thiruvananthapuram, Kerala 695581, India

Introduction

L-asparaginase (E.C. 3.5.1.1), also known as L-asparagine amidohydrolase, catalyzes the hydrolysis of asparagine to aspartic acid and ammonia. It is an important chemotherapeutic drug used for the treatment of acute lymphoblastic leukemia (ALL) (Nguyen et al. 2018). It is also used for the treatment of other malignancies such as acute myelocytic leukemia, acute myelomonocytic leukemia, chronic lymphocytic leukemia, lymphosarcoma, melanosarcoma, non-Hodgkin's lymphoma, pancreatic carcinoma, and reticulosarcoma (Kidd 1953; Broome 1963; Izadpanah Qeshmi et al. 2018; Van Trimpont et al. 2022). L-asparaginase (L-ASNase) formulations contribute to one-third of the global requirement of leukemia and lymphoma therapeutic agents and meet 40% of the global enzyme demands (Warangkar and Khobragade 2010). It is listed on the World Health Organization's list of essential medicines, which catalogues the effective and safe medicines required by the health system (WHO model list of essential medicines 2019).

L-ASNase also finds application in the reduction of acrylamide formation in the food processing industry. The International Agency for Research on Cancer (IARC) lists acrylamide as a group B2 agent that is probably carcinogenic to humans (IARC 1994). The association between acrylamide and cancers, viz. breast, ovarian, endometrial, and renal cancers, has been the subject of several investigations (Hogervorst et al. 2007; Kumar et al. 2018a). Acrylamide is formed during baking or frying of starchy foods at 120 °C or above, due to the reaction between α -amino group of L-asparagine and the carbonyl groups of reducing sugars (Maillard reaction) (Stadler et al. 2004). Pre-treatment of such foods with L-ASNase reduces the availability of the precursor, L-asparagine and prevents acrylamide formation. About 88 to 90% reduction of acrylamide has been achieved with the addition of ASNase, without compromising the taste/appearance of the end product (Jia et al. 2021; Lineback et al. 2012; Medeiros Vinci et al. 2012; Cachumba et al. 2016). Thus, L-ASNase serves as a two-pronged fork with respect to its application in the prevention and treatment of cancer.

L-asparaginases show a widespread distribution across all the three domains of life. However, bacterial L-ASNases from *Erwinia chrysanthemi* and *Escherichia coli* have been in use for the treatment of ALL and other malignancies for more than 30 years. As the use of wild strains often provides limited yields and poses challenges in downstream processing, mutant bacterial strains such as *E. coli* are used for rapid and easy overexpression of recombinant L-ASNases (Castro et al. 2021; Pokrovskaya et al. 2012a, b; Wang et al. 2021b). Nonetheless, the usage of current ASNase formulations

is associated with a plethora of severe side effects such as central nervous system toxicity, liver and pancreas dysfunction, fever, skin rashes, edema, hyperglycemia, leucopenia, hemorrhage, hypersensitivity reactions, and anaphylactic shock, chiefly on account of their intrinsic glutaminase activity (Ramya et al. 2012; Batool et al. 2016; Izadpanah Qeshmi et al. 2018). This has led to an accelerated interest in the search for novel L-asparaginases with good therapeutic potential and minimum side effects.

The present study reports the cloning and expression of L-asparaginase contig obtained from whole metagenome shotgun sequencing of DNA from *Sardinella longiceps* gut microbiota in *E. coli*. The overexpressed L-ASNase was purified to homogeneity and subjected to physicochemical and kinetic characterization. Further, the protein was investigated for cytotoxicity against K-562 (chronic myeloid leukemia) and MCF-7 (breast cancer) cell lines and for acrylamide mitigation in bread and biscuits.

Materials and methods

Sequence analysis and cloning of L-asparaginase contig SlcA

Metagenomic DNA extracted from *Sardinella longiceps* was subjected to whole metagenome shotgun sequencing on the Illumina HiSeq 2000 platform (NCBI accession SRR6677342). The de novo assembly of reads into contigs was performed using Ray Meta assembler (ver 2.3.1) (Boisvert et al. 2012). The prediction of the open reading frame was performed using MetaGeneAnnotator (MGA) (Noguchi et al. 2008). The predicted ORFs were searched against the NCBI nonredundant (nr) protein database using DIAMOND (ver 0.7.9.58) for functional annotation (Buchfink et al. 2015). From 4,271 reads, five potential L-asparaginase contigs were identified. L-asparaginase contig, designated as SlcA, was chosen for cloning and expression studies on the basis of in silico experiments (data not shown). The nucleic acid sequence of 1,014 bp contig SlcA (Accession No. MT740290) was compared against other sequences in GenBank using 'blastn' tool of NCBI (<https://blast.ncbi.nlm.nih.gov>) (Altschul et al. 1997). Further, the sequence was translated into amino acid sequence using 'translate' tool of ExPASy server (<https://web.expasy.org/translate/>) (Gasteiger et al. 2003), and the deduced amino acid sequence was used as a query for protein-BLAST (blastp) against the nonredundant protein database of NCBI (<https://blast.ncbi.nlm.nih.gov>) (Altschul et al. 1997).

Multiple sequence alignment of the amino acid sequence of SlcA with other reported L-asparaginases

was performed using ClustalW tool of EMBL-EBI (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Larkin et al. 2007). As bacterial L-asparaginases are classified into type I and type II ASNases based on subcellular location, sequence similarity, and substrate affinities (Izadpanah Qeshmi et al. 2018), both types of asparaginases were selected for multiple sequence alignment. The sequences chosen for alignment were type II L-ASNases from *E. coli* strain K-12, *Erwinia chrysanthemi*, and *Mycobacterium tuberculosis* and type I L-ASNases from *E. coli*, *Photobacterium* sp. J15, *Pyrococcus horikoshii*, *Pyrococcus furiosus* DSM 3638, *Vibrio campbellii*, and *Yersinia pseudotuberculosis*. Phylogenetic analysis was conducted to infer the evolutionary relatedness of the L-ASNases reported from various sources. MEGA-X was used for the construction of phylogenetic tree (Kumar et al. 2018b).

Cloning of L-asparaginase contig SlcA

PCR amplification of SlcA contig was carried out using the primers Asn2-FR (5'-GCG GGA TCC ATG GAA AGA AAA CAT ATT-3') and Asn2-RR (5'-ACT CTC GAG TTA CAA GGT TAA CTC ACC-3') with BamHI and XhoI restriction sites, respectively for directional cloning into pET 32a(+) vector. The amplicons were subjected to purification, double digestion and ligation to pET 32a(+) vector, designed for the creation of fusion proteins with a Trx-His-S tag. The recombinant vector was transformed into *E. coli* BL21(DE3) pLysS cells for overexpression of the protein, hereafter named as SlpA.

Overexpression and purification of recombinant L-asparaginase SlpA

The expression of SlpA was induced by IPTG induction (1 mM) of the cells at 30 °C. The culture was harvested by centrifugation, and the cell pellet was resuspended in ice-cold Tris-HCl buffer (50 mM), pH 8, containing 1× protease inhibitor cocktail (Sigma-Aldrich, India). The cell pellet was disrupted by sonication in an ice bath at 40% amplitude for 10 min (15 s on, 45 s off) (Sonics Vibra-Cell VCX 500 ultrasonic processor). The cell-free supernatant (crude SlpA) was subjected to purification using Ni-NTA spin columns (Qiagen, Germany) following the manufacturer's protocols. The eluted protein was dialyzed against 50 mM Tris-HCl buffer, pH 8, and was checked for enzyme activity. Even though purified SlpA recorded a protein content of 2.4 mg/mL, it did not show any asparaginase activity. Crude SlpA was therefore subjected to

ammonium sulfate precipitation, gel filtration, and ion exchange chromatography for purification.

Briefly, the cell-free supernatant obtained after sonication was concentrated by precipitation using 30–60% ammonium sulphate. The precipitate was dissolved in 50 mM Tris-HCl buffer (pH 8) and dialyzed against the same buffer to remove ammonium sulphate. SlpA was further purified by gel filtration chromatography on Sephadex G-75 column (Sigma-Aldrich, India). Final purification of SlpA was done by anion exchange chromatography on DEAE cellulose column using gradient concentrations of NaCl (0–1 M) prepared in Tris-HCl buffer (pH 8). The elution fractions showing asparaginase activity were dialyzed, pooled, and concentrated using 10 kDa Amicon filter (Merck Millipore, Germany).

Assay of enzyme activity

L-asparaginase activity was measured according to the method of Imada et al. (1973). The method uses the nesslerization process to measure ammonia liberated by the hydrolysis of L-asparagine. The reaction mixture comprised 100 mM L-asparagine in 50 mM Tris-HCl buffer (pH 8.0) and suitably diluted SlpA. Following incubation at 37 °C for 30 min, the reaction was terminated by the addition of 1.5 M trichloroacetic acid. The reaction mixtures were clarified by centrifugation, and ammonia liberated in the reaction was measured by the addition of Nessler's reagent. The tubes were incubated at room temperature for 15 min, and absorbance was measured at 425 nm. One unit (U) of L-asparaginase activity is defined as the amount of enzyme that liberates 1 µmol of ammonia per minute at pH 8 and 37 °C.

Determination of L-glutaminase activity was performed by incubation of the purified SlpA with 100 mM L-glutamine prepared in 50 mM Tris-HCl buffer (pH 8). The enzyme reaction was performed by the nesslerization process as described above.

Determination of molecular weight and oligomeric state of SlpA

The purity of the protein fractions collected at different stages of purification was analyzed using SDS-PAGE. The protein fractions were mixed with SDS sample buffer, heated at 95 °C for 10 min, and resolved in 12% SDS-PAGE gel. After electrophoresis, the gel was stained with Coomassie brilliant blue R-250 for 1 h and destained for visualization of the protein bands. Following all rounds of purification, the histidine tag of the protein was removed by thrombin digestion for the determination of actual molecular weight of the protein. The oligomeric state of the protein was determined

using native PAGE. The gel images were captured using ChemiDoc™ XRS system (BioRad, USA).

MALDI-TOF spectrometry

MALDI-TOF analysis for protein identification was performed at the Proteomics Facility, Rajiv Gandhi Centre for Biotechnology, Trivandrum, Kerala. The protein band corresponding to SlpA was excised from the SDS-PAGE gel and was subjected to in-gel reduction and alkylation, followed by trypsin digestion. The extracted peptides were dried to remove solvent traces, dissolved in water, and co-crystallized with α -cyano-4-hydroxycinnamic acid matrix on the target plate (384-well ground steel plate, Bruker Daltonics, Germany), and external peptide mass calibration was applied (peptide mixture-1, Bruker) according to manufacturer's instructions. MALDI-TOF/MS/MS analysis was performed on UltrafleXtreme (Bruker Daltonics) in reflectron positive ion mode using 19-kV acceleration voltage. The peptide mass fingerprint (PMF) data was acquired in the mass range of 700–3500 m/z. To confirm the identity of the protein, the MS/MS data from LIFT TOF/TOF spectra were combined with the corresponding MS peptide mass (PMF), and mass spectra were imported into the database search engine (BioTools v2.2 connected to MASCOT, version 2.2.04; Matrix Science).

Biochemical characterization

The optimum pH for SlpA activity was determined by conducting an enzyme assay at different pH in the range from 1 to 13. L-asparagine substrate was prepared in buffer systems of different pH, viz. hydrochloric acid/potassium chloride buffer (pH 1–2), citric acid/sodium citrate buffer (pH 3–6), Tris-amino methane/hydrochloric acid buffer (pH 7–9), sodium carbonate/sodium bicarbonate buffer (pH 10–11), and sodium hydroxide/potassium chloride buffer (pH 11–12).

The optimum temperature for SlpA activity was determined by assaying enzyme activity at 10, 20, 25, 30, 37, 40, 50, and 60 °C for 30 min. The thermal stability of SlpA was determined by pre-incubating the enzyme for 1 h (10–60-min intervals) at different temperatures ranging from 10 to 50 °C.

Determination of kinetic parameters

Kinetic studies of the purified SlpA (50 U/mg) were conducted to determine K_m and V_{max} . K_m is the substrate concentration at which the reaction velocity is half maximum,

and V_{max} is the maximum velocity of the enzyme reaction. The enzyme assay was conducted by varying L-asparagine substrate [S] concentrations from 10 to 100 mM, and enzyme activity [V] was calculated. Kinetic parameters, K_m (mM) and V_{max} (U/mL), were calculated from the Lineweaver–Burk plot.

Anticancer studies of SlpA

In vitro cytotoxicity assay of SlpA against cancer cell lines

The anticancer activity of purified SlpA was investigated on K-562 (chronic myeloid leukemia) and MCF-7 (breast cancer) cell lines. K-562 cell line was propagated in RPMI 1640 medium supplemented with 10% FBS (fetal bovine serum) and penicillin–streptomycin solution (100 IU/mL). MCF-7 cell line was propagated in DMEM medium supplemented with 10% FBS (fetal bovine serum) and penicillin–streptomycin solution (100 IU/mL).

For MTT assay, the cells were seeded at a density of 5000 cells per well in a 96-well plate and incubated for 24 h at 37 °C in a 5% CO₂ incubator. The wells were treated with different concentrations of purified SlpA and incubated for 24 h at 37 °C in a 5% CO₂ incubator. After incubation, the media was removed, and a 100 μ L working solution of MTT dye (1 mg/mL) prepared in the respective media was added and incubated for 2 h. The MTT dye was removed, and the formazan crystals were dissolved using 100 μ L DMSO. The absorbance was measured at 570 nm using a Tecan multimode plate reader, and the cell viability [proliferation rate (PR)] was calculated using the formula:

$$PR = \frac{\text{Absorbance of test} - \text{Absorbance of blank}}{\text{Absorbance of control} - \text{Absorbance of blank}} \times 100$$

Morphological evaluation of apoptosis

Acridine orange/ethidium bromide (AO/EB) staining

The cells were seeded at a density of 5000 cells per well in 96-well with a flat bottom and incubated for 24 h at 37 °C in a 5% CO₂ incubator. The cells were treated with IC₅₀ concentration of SlpA and incubated for 24 h at 37 °C in a 5% CO₂ incubator. After incubation, the media was discarded, and 25 μ L of acridine orange (7.5 μ g)/ethidium bromide (25 μ g) stain was added, mixed well, and observed under a fluorescent microscope with a FITC filter (McGahon et al. 1995; Kasibhatla et al. 2006).

Scanning electron microscopy

The cells were grown in a T25 culture flask for 24 h and treated with SlpA for 24 h. After the incubation period, cells were centrifuged at 2000 rpm for 10 min. The cell pellet was collected and rinsed twice with fresh PBS buffer. The buffer solution was replaced with ethanol solution in dehydration series –50%, 60%, 70%, 80%, and 90%, leaving each step for 15–20 min. Finally, the samples were dehydrated with 100% ethanol twice. Scanning electron microscopy was performed on Jeol JSM-6390 LV scanning electron microscope at Sophisticated Test Instrumentation Centre, CUSAT, Cochin, Kerala.

Toxicological evaluation of SlpA against normal cells

In vitro cytotoxicity assay against normal human lymphocytes

Whole blood was collected and lymphocytes were isolated by Ficoll density gradient centrifugation. The isolated lymphocytes were seeded at a density of 5000 cells per well in a 96-well plate with a flat bottom and incubated for 24 h at 37 °C in a 5% CO₂ incubator. Phytohemagglutinin was added to each well, and the wells were treated with different concentrations of SlpA for 48 h, and MTT assay was performed for evaluation of cytotoxicity.

In vitro hemolysis assay

The hemolytic effect of crude and purified SlpA was checked on human blood (erythrocytes) using the blood agar plate method. The crude and purified SlpA and a phosphate buffer control were added into separate wells punched on the blood agar plates and were incubated at 37 °C for 24 h. A translucent zone of clearance following incubation is indicative of hemolysis.

For quantitative hemolytic assay, heparinized human blood cells (erythrocytes) were washed thrice with an equal volume of 150 mM NaCl and were suspended in 100 mM sodium phosphate buffer at pH 7.4. The cells were incubated with different concentrations of purified SlpA (2, 4, 6, 8, 10, 20, 30, 40 U/mL) for 24 h at 37 °C. The cells were centrifuged at 2500 rpm for 15 min and the pellet was discarded. The optical density of the supernatant was measured at 541 nm. Sodium phosphate buffer (pH 7.4) incubated with the erythrocytes served as the blank, and the blood cells incubated with 1% Triton X-100 served as the positive control.

Acrylamide mitigation potential of SlpA

Preparation of biscuit and bread

Baking of biscuits was performed according to the method described by Anese et al. with minor modifications (Anese et al. 2011). Briefly, the short dough formulation for a single biscuit consisted of cake flour (10 g), sucrose (2.5 g), salt (0.12 g), shortening (1.6 g), glucose (0.13 g), sodium bicarbonate (0.05 g), ammonium bicarbonate (0.1 g), and water (2.5 mL). The glucose, sodium bicarbonate, and ammonium bicarbonate were dissolved in water, and different concentrations of SlpA ranging from 0.5 U/g to 100 U/g flour were added. The remaining ingredients were mixed, and the dough was kneaded and held for 30 min at 30 °C. The dough was sheeted into 0.3-cm-thick, 7-cm-diameter circles and was kept at 30 °C for another 30 min. The samples were baked at 260 °C for 10 min.

Bread preparation was done according to the method of Jiang et al. (2005) with minor modifications (Jiang et al. 2005). The dough formulation for each loaf of bread consisted of wheat flour (50 g), dry *Saccharomyces cerevisiae* yeast (1.0 g), sucrose (2.5 g), salt (0.75 g), vegetable oil (1.5 g), and water (30 mL). Purified SlpA ranging from 0.5 to 100 U/g flour were dissolved in water and incorporated into the flour. The dough was kneaded and held at 30 °C for 30 min. The dough was then rolled out and kept at 30 °C for 30 min for its proofing to three times its volume. The loaves were baked at 200 °C for 15 min. The dough pH of the samples was tested in a 10% w/v suspension of dough sample prepared in Milli-Q water.

Determination of acrylamide content

By shaking incubation in a water bath at 30 °C for 30 min, 1 g each of biscuit and bread samples were defatted with 10 mL of hexane. The hexane layer was removed, and 10-mL acetonitrile was added to the sample and incubated in the water bath for 2 h with agitation. The samples were centrifuged, and the acetonitrile fractions were collected. The fractions were clarified by filtration and concentrated by vacuum evaporation. This was used for LC–MS analysis. Acrylamide solutions prepared in the concentration range of 0.5 to 2.5 ppm were used as standards for calibration.

LC–MS analysis was performed at CARE KERALAM Ltd, Koratty, Thrissur, Kerala. LC–MS analysis was

performed on an Agilent 1260 Infinity High-Performance Liquid Chromatography (HPLC) system (Santa Clara, CA, USA), coupled to Agilent 6120 single quadrupole mass spectrometric detector equipped with positive electrospray ionization interface. The analytical column was Agilent Eclipse plus C-18 (4.6 × 250 mm). The mobile phase was acetonitrile: 0.1% acetic acid in water (60:40 v/v) with a flow rate of 0.4 mL/min. Data acquisition was performed, with a run time of 8 min in selected ion-monitoring (SIM) mode ($[M+H]^+$ ion peak at m/z 72) using the following interface parameters: a drying gas (N_2) flow of 10 L/min, nebulizer pressure of 50 psi, drying gas temperatures of 350 °C, and a capillary voltage of 4200 V.

Statistical analysis

All the experiments were repeated thrice. The results were expressed as mean ± standard deviation, unless indicated otherwise. Statistical analysis was performed using GraphPad Prism (ver.6.01). One-way ANOVA with Tukey post hoc test was used to compare the means, and p values < 0.05 were considered significant.

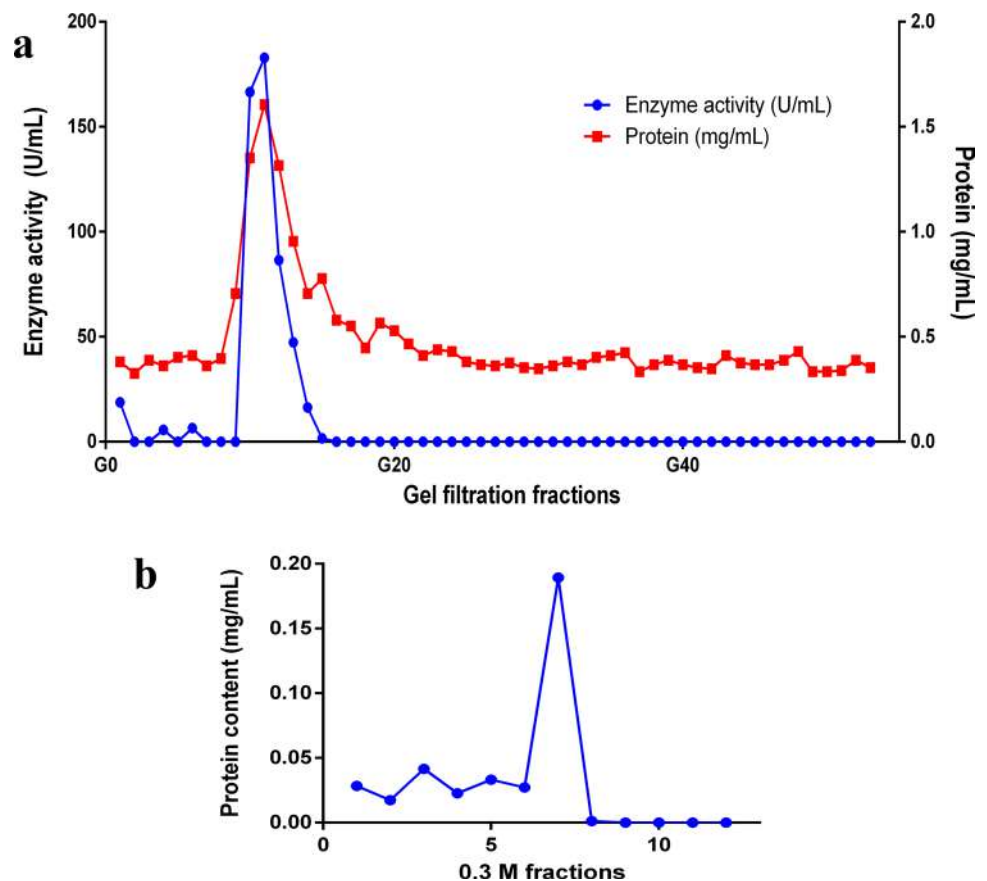
Results

Sequence analysis of contig SlcA

Nucleotide BLAST search using the MegaBlast algorithm showed 77.73% identity of the SlcA contig to chromosome 1 sequence of *Vibrio metoecus* strain 08–2459. SlcA sequence also showed identity to chromosome sequences of several *Vibrio* sp. (Table S1). Protein BLAST analysis of the deduced amino acid sequence of SlcA (Fig S1) revealed 98.81% identity to L-asparaginase from *Photobacterium leiognathi*. The remaining hits also showed identity to L-asparaginases from several *Photobacterium* sp., thereby confirming the identity of the contig (Table S2). The dearth of studies on L-asparaginases related to SlcA warrants detailed investigation of the SlcA contig and the encoded protein, SlpA.

Multiple sequence alignment showed that 6.2% of the residues of SlpA were conserved across the examined L-asparaginases (Fig S2). Phylogenetic analysis construed significant homology between SlpA and type I L-ASNases (Fig S3). Among all the type I L-asparaginases considered, SlpA was more related to L-ASNases

Fig. 1 Gel filtration and ion exchange chromatographic profiles of L-asparaginase SlpA. **a** Gel filtration profile. SlpA was eluted between 10th and 15th fractions. **b** Ion exchange chromatographic profile. SlpA was eluted in the 7th fraction



from *Photobacterium* sp. J15 (pairwise distance 0.08) and *V. campbellii* (pairwise distance 0.17). SlpA showed relatedness to L-ASNases from *E. coli* (type I) (pairwise distance 0.31) and *Y. pseudotuberculosis* (pairwise distance 0.33) as well, but formed a separate clade. Type I L-ASNases formed a distinct cluster in the analysis with respect to the FDA-approved type II L-ASNases belonging to *E. coli* (marketed as Elspar), *E. chrysanthemi* (marketed as Erwinaze), and L-ASNase of *M. tuberculosis*. Type III L-ASNase from *E. coli*, used as an outgroup clustered separately from type I and II L-ASNases, with a pairwise distance of 2.17 from SlpA.

Expression and purification of L-asparaginase SlpA

Crude L-asparaginase SlpA (cell-free supernatant) obtained after sonication recorded asparaginase activity of 71.08 ± 5.84 U/mL with a specific activity of 45.87 U/mg. SlpA purified by immunoaffinity chromatography (Ni-NTA columns) did not show any L-asparaginase activity. This might be the result of denaturation or incorrect folding of the protein during purification. The enzyme activity and specific activity of SlpA after ammonium sulphate precipitation and dialysis were found to be 307.71 ± 3.14 U/mL and 104.36 U/mg, respectively. This purification step resulted in a 2.28-fold increase of protein

purity. SlpA was further purified by gel filtration chromatography on Sephadex G-75 column. The elution profile showed that SlpA was eluted in the fractions 10 to 15 (Fig. 1a). Enzyme activity and specific activity of SlpA after gel filtration chromatography were 182.93 ± 3.23 U/mL and 227.81 U/mg, respectively. These active fractions were pooled and used for the next round of purification by ion exchange chromatography. SlpA was eluted in the 0.3 M NaCl fraction (Fig. 1b) and was purified to homogeneity with a specific activity of 478.81 U/mg and purification fold of 10.44. Purified SlpA was also assayed for glutaminase activity. SlpA did not exhibit any intrinsic glutaminase activity.

Determination of molecular weight and oligomeric state of SlpA

SDS-PAGE analysis of purified SlpA depicted a single distinctive protein band with an approximate subunit molecular weight of 52 kDa (Fig. 2a). After the removal of Trx-His-S tag by thrombin digestion, the molecular weight of the SlpA was 37 kDa (Fig. 2b). Native-PAGE was performed to assess the oligomeric state of the purified protein. Purified SlpA showed an approximate molecular weight of 74 kDa on the native PAGE gel (Fig. 2c), indicating its homodimeric nature.

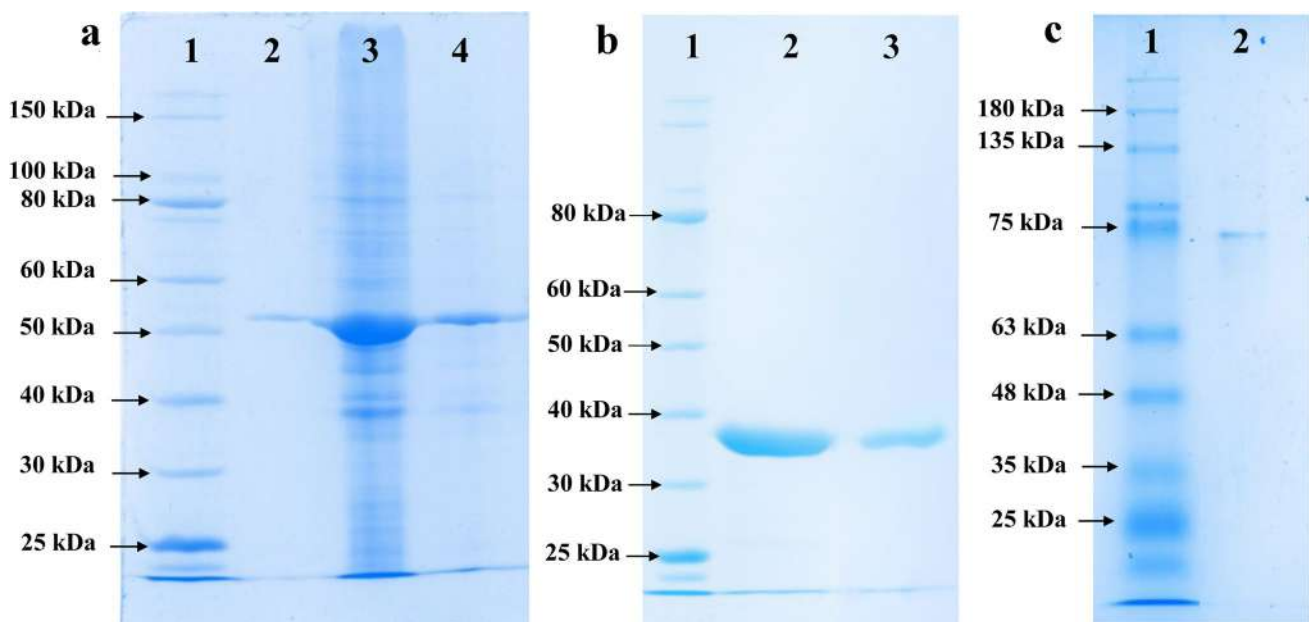


Fig. 2 Electrophoretic analyses of L-asparaginase SlpA. **a** SDS-PAGE profile of purification steps of SlpA. Lane 1, NEB marker; lane 2, purified SlpA (DEAE cellulose eluate); lane 3, ammonium sulphate precipitate (30–60%) fraction; lane 4, gel filtration fraction. **b** SDS-PAGE profile of thrombin digested SlpA. Lane 1, NEB marker;

lane 2, purified SlpA (DEAE cellulose eluate) after thrombin digestion; lane 3, Ni-NTA purified SlpA after thrombin digestion. **c** Native PAGE analysis. Lane 1, origin broad range protein marker; lane 2, purified SlpA after thrombin digestion

Fig. 3 Peptide fingerprinting map obtained by MALDI-TOF-MS/MS of L-asparaginase SlpA

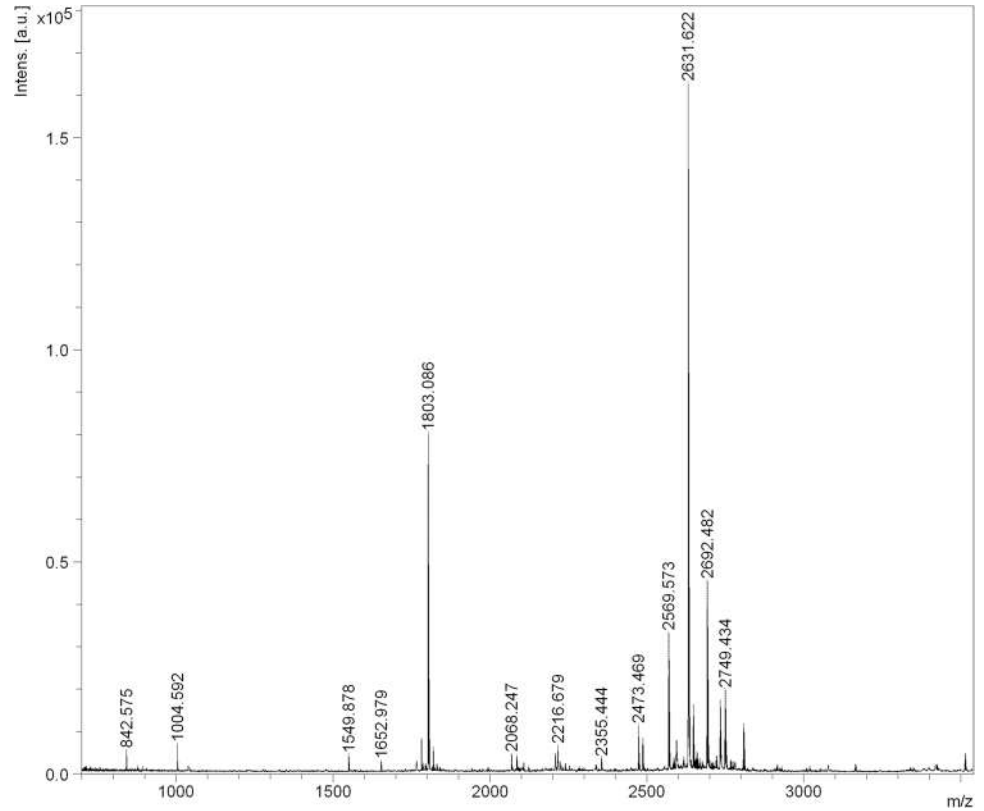
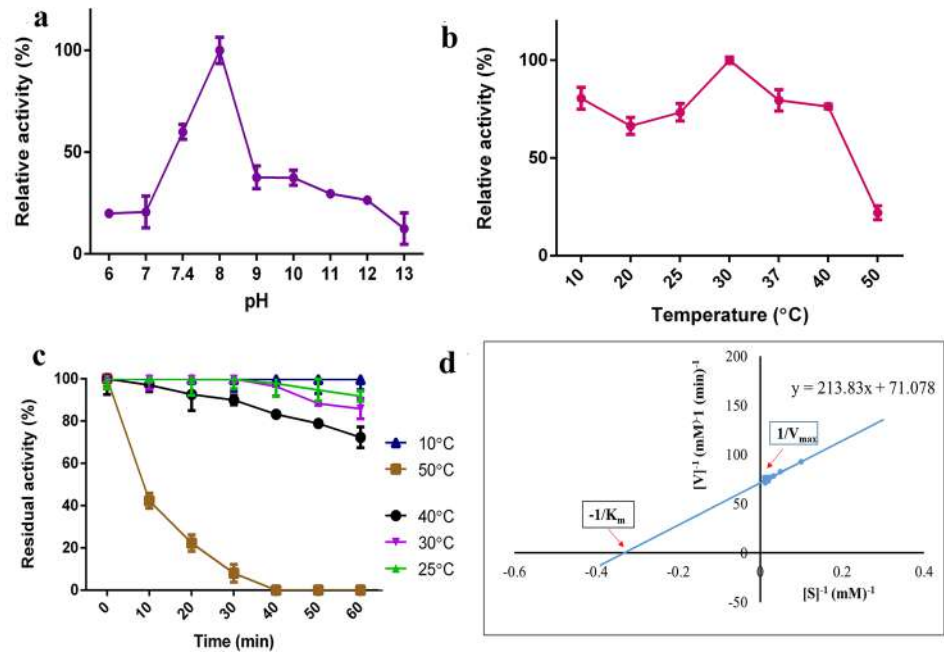


Fig. 4 Biochemical characterization of L-asparaginase SlpA. **a** Effect of pH on SlpA activity. **b** Effect of temperature on SlpA activity. **c** Effect of temperature on the stability of SlpA. **d** Lineweaver Burk plot of SlpA



MALDI-TOF spectrometry

MALDI-TOF/MS/MS analysis revealed the presence of two major peaks (2631.622 m/z; 1803.086 m/z) (Fig. 3). These peaks showing maximum abundance were sequenced, and the resultant peptide sequence was investigated across

MASCOT search engine using the database for *Photobacterium* downloaded from UniProt. The peptide peak at 2631.622 m/z showed similarity to L-ASNase from *Photobacterium aphoticum* with a protein score of 95, sequence coverage of 7%, and nominal mass of 35,671 (Figure S4). The second major peak at 1803.086 m/z showed similarity

to L-ASNase from *Photobacterium leiognathi* with a protein score of 71, sequence coverage of 4%, and nominal mass of 37,288 (Figure S5). The protein scores of 95 and 71 obtained for the peptides from MALDI-TOF-MS/MS indicated significant identity of the peptides to amino acid sequence of L-ASNase from *Photobacterium* sp. The molecular weight deduced from MASCOT search confirmed the subunit molecular weight of SlpA determined from SDS-PAGE analysis.

Biochemical characterization and kinetic analysis of SlpA

The enzyme activity of SlpA increased as the pH increased from 6 to the alkaline pH of 8 but declined with further pH increase. Thus, SlpA was active in the alkaline pH range with a pH optimum at 8. The enzyme activity was lost below pH 6. However, SlpA possessed good activity at the physiological pH of 7.4, indicating its suitability for application in cancer treatment (Fig. 4a).

The activity of SlpA increased with an increase in temperature up to 30 °C and declined gradually. From the temperature profile depicted in Fig. 4b, it follows that the optimum temperature for enzyme activity is 30 °C. The enzyme showed 80% activity at a low temperature of 10 °C as well as at 37 °C. There was a complete loss of enzyme activity at 60 °C. This implied that SlpA is best active at low temperatures. Temperature stability studies (Fig. 4c) indicate that SlpA is moderately stable up to 40 °C. The enzyme showed more than 70% residual activity after 1 h incubation at 40 °C. But asparaginase activity decreased to 8% after a 30-min incubation at 50 °C. The stability of SlpA up to 40 °C is encouraging to the application and development of the enzyme as an anticancer drug, as well as its use as a food processing

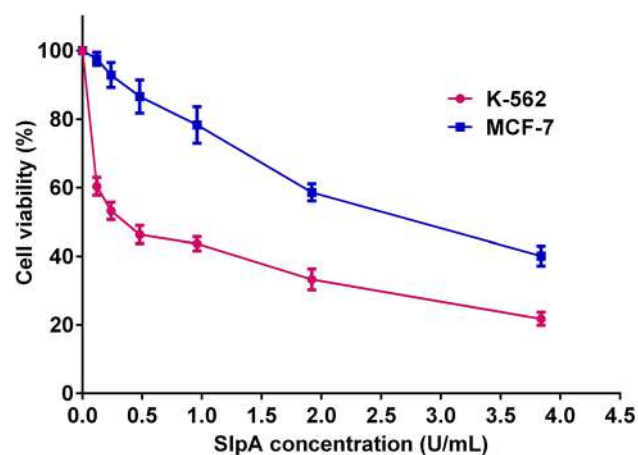


Fig. 5 Cytotoxicity assay of L-asparaginase SlpA against cancer cell lines

aid in the baking industry where kneading and dough proofing are usually carried out at temperatures ranging up to 35 °C (Nguyen et al. 2017; Rosell and Collar, 2009).

SlpA showed a hyperbolic response to increasing concentrations of the asparagine, well in accordance with the Michaelis–Menten equation. Kinetic parameters, K_m and V_{max} of the purified L-asparaginase SlpA were found to be 3.008 mM asparagine and 0.014 mM/min, respectively (Fig. 4d).

Anticancer studies of SlpA

In vitro cytotoxicity assay of SlpA against cancer cell lines

L-asparaginase SlpA affected the viability of K-562 and MCF-7 cell lines in a dose-dependent manner (Fig. 5). A concentration of 0.12 U/mL concentration of SlpA reduced the viability of K-562 and MCF-7 cell lines to 60% (40% loss of viability) and 98% (2% loss of viability), respectively. The viability of K-562 and MCF-7 cell lines decreased to 46% and 87% upon treatment with 0.48 U/mL concentration of SlpA. The K-562 viability decreased to 22% at 3.84 U/mL concentration of SlpA. The same concentration of SlpA also reduced the viability of MCF-7 cell line, but only to 40%. SlpA showed significant cytotoxic activity against K-562 cell line at concentrations ranging from 0.12 to 3.84 U/mL (p value < 0.05). In the case of MCF-7 cell line, no significant cytotoxicity was recorded at SlpA concentrations of 0.12 and 0.24 U/mL. However, SlpA concentrations ranging from 0.48 U/mL to 3.84 U/mL showed significant cytotoxicity against MCF-7 cell line (p value < 0.05). The IC_{50} values recorded for K-562 and MCF-7 cell lines were 0.3443 and 2.692 U/mL, respectively. These values indicate that the effectiveness of SlpA differs between cell types and that it shows better cytotoxic activity against K-562 cell line (chronic myeloid leukemia) compared to MCF-7 cell line (breast cancer).

The cytotoxicity induced by purified SlpA was confirmed by assessing the morphological changes in K-562 and MCF-7 cells treated with IC_{50} concentrations of SlpA for 24 h using acridine orange/ethidium bromide (AO/EB) staining (Fig. 6). All the cells in the control group showed normal green-colored nuclei, indicative of viable cells. SlpA-treated cell populations of both K-562 and MCF-7 cell lines possessed a mixture of viable and nonviable cells in various stages of apoptosis. SlpA-treated K-562 cells showed orange nuclei with patches of orange color indicating condensed chromatin, indicative of late apoptosis. Some of the K-562 cells with green nuclei also showed condensed chromatin patches, signifying early apoptosis. SlpA-treated MCF-7 cells showed

Fig. 6 Morphological evaluation of apoptosis after AO/EB double staining. **a** Control K-562 cells. **b** SlpA-treated K-562 cells. **c** Control MCF-7 cells. **d** SlpA-treated MCF-7 cells

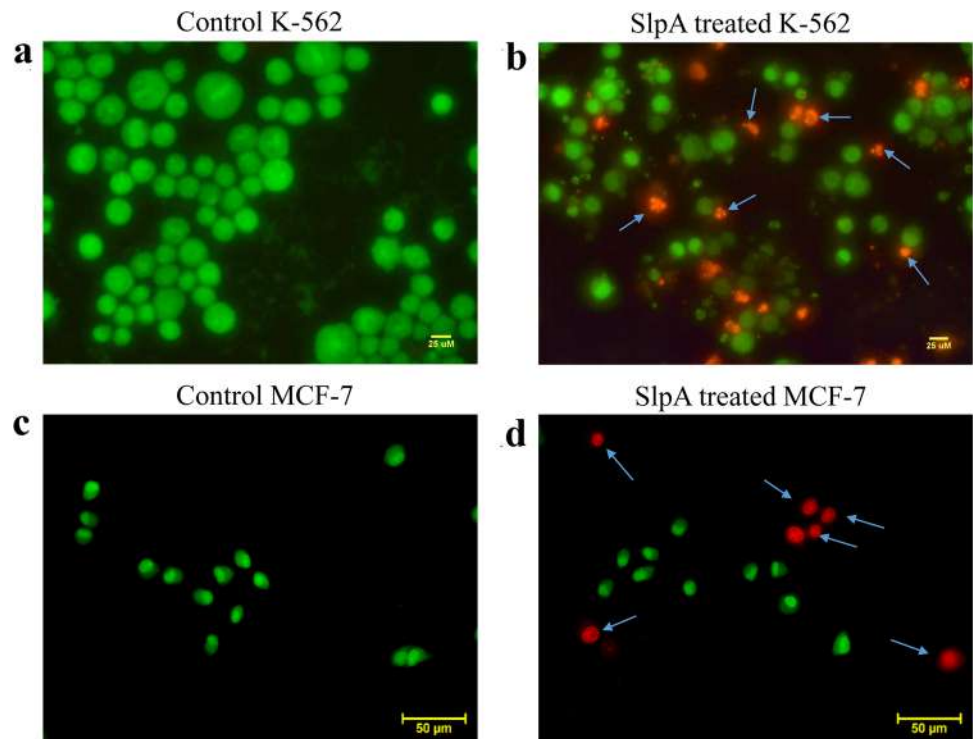
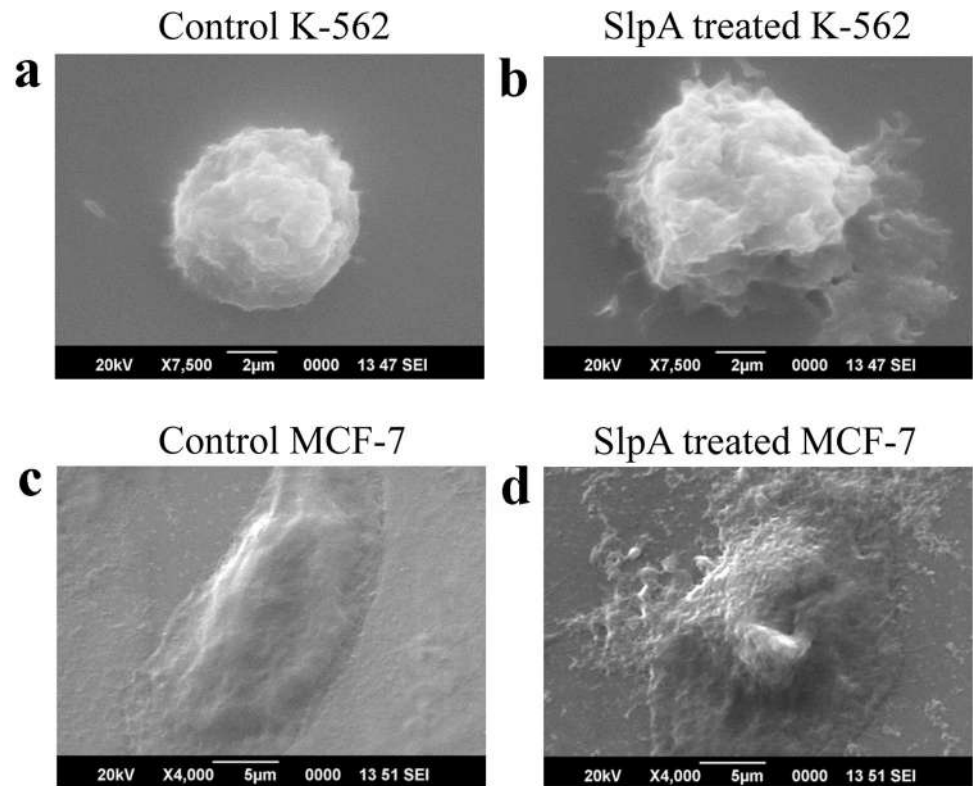


Fig. 7 SEM micrographs of control (untreated) and SlpA-treated K-562 and MCF-7 cell lines. **a** Control K-562 cells. **b** SlpA-treated K-562 cells. **c** Control MCF-7 cells. **d** SlpA-treated MCF-7 cells



orange-colored nuclei with some orange patches of condensed chromatin, representing late apoptotic cells.

Upon SEM analysis, control (non-treated) K-562 cells were observed to be round in shape with a relatively smooth

surface and small numbers of short microvilli (Fig. 7). SlpA-treated K-562 cells lost their near perfect round shape and showed cell shrinkage, membrane blebbing, and cellular fragmentation, indicative of cell death. The control MCF-7

Fig. 8 Toxicological evaluation of L-asparaginase SlpA. **a** Cytotoxicity assay of SlpA against normal human lymphocytes. **b** In vitro hemolytic assay on blood agar plate. 1. Tris-HCl buffer (control) 2. Crude SlpA 3. Purified SlpA (20 U). **c** Quantitative measurement of hemolytic activity. The first 8 tubes (from left) have 2, 4, 6, 8, 10, 20, 30, and 40 U/mL of purified SlpA. Negative control: Tris-HCl buffer. Positive control: 1% Triton X-100

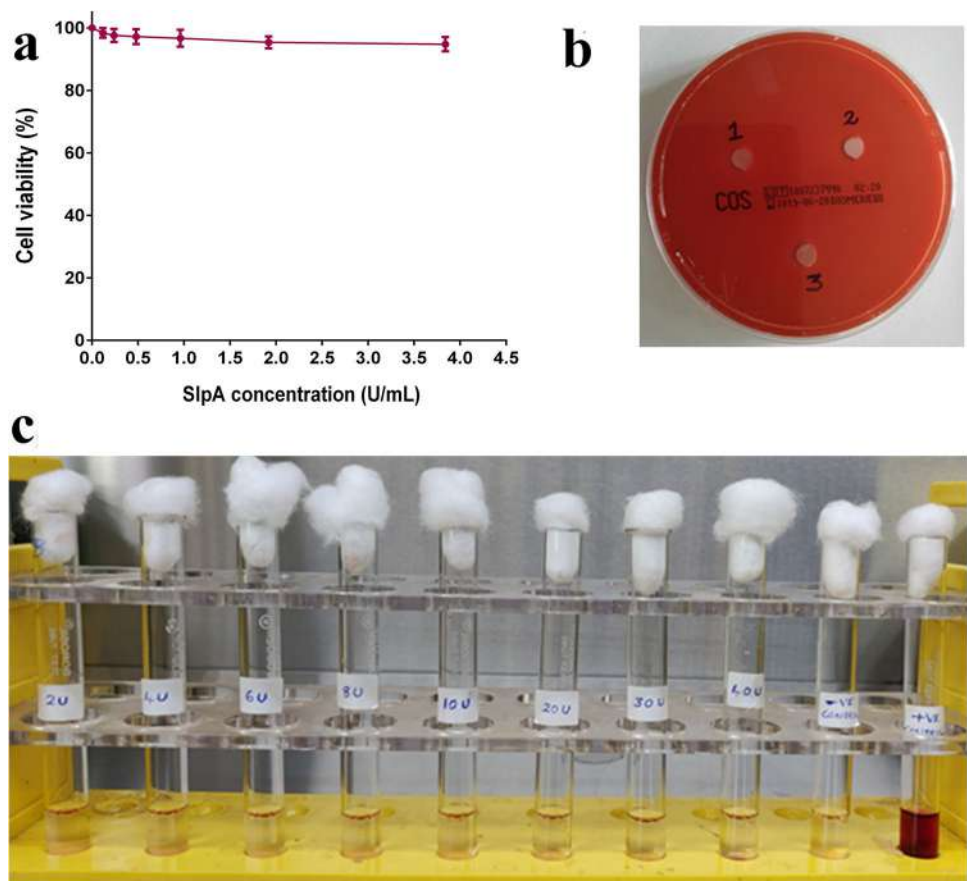
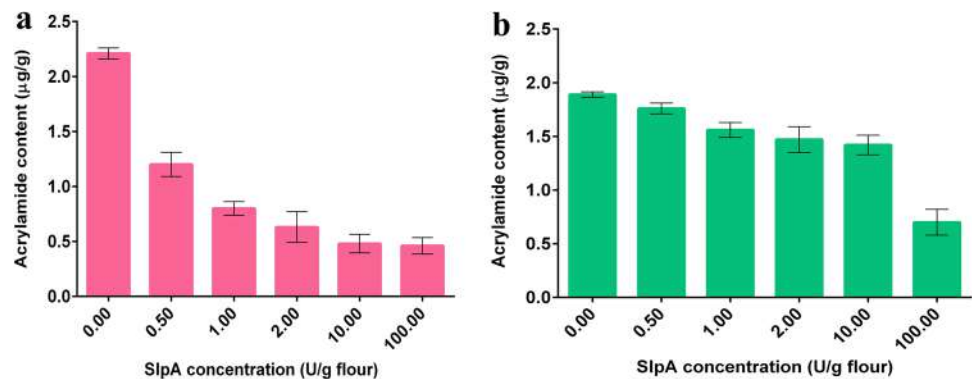


Fig. 9 Reduction of acrylamide formation in biscuits and bread as a function of SlpA dosage. **a** Acrylamide content in SlpA-treated biscuit samples. **b** Acrylamide content in SlpA-treated bread samples



cells revealed the presence of a semi-flattened surface structure containing microvilli with few extended lamellipodia called membrane ruffles. SlpA-treated cells appeared round shaped and shrunken, indicative of lamellipodia retraction. Convuluted membrane and membrane blebbing were also visible in the treated cells. Membrane blebbing is an important feature of apoptosis that occurs due to extensive cytoskeletal rearrangement that alters cell shape and distribution of organelles.

Toxicological evaluation of SlpA against normal cells

SlpA did not cause a significant loss of viability of normal lymphocytes at the IC₅₀ concentrations for K-562 (0.3443 U/mL) and MCF-7 (2.692 U/mL) cell lines (*p* value > 0.05). More than 94% of lymphocytes remained viable even after 48-h treatment with 3.84 U/mL concentration of SlpA (Fig. 8a).

In vitro hemolytic activity of purified SlpA was checked using the blood agar plate method (Fig. 8b). No zone of

hemolysis was observed even with 20 U of purified SlpA. Quantitative hemolytic assay did not detect any signs of hemolysis at SlpA concentrations ranging from 2 to 40 U/mL (Fig. 8c).

Acrylamide mitigation potential of SlpA

Addition of SlpA resulted in a clear, dose-dependent reduction of acrylamide content in the biscuit sample (Fig. 9a). The acrylamide content in the untreated biscuit sample was 2.21 µg/g. Pre-treatment of biscuit dough with SlpA concentrations ranging from 0.5 to 100 U/g caused a significant reduction of acrylamide content, compared to control (p value < 0.05). Treatment of dough with 0.5 to 100 U/g flour of SlpA reduced the acrylamide levels from 1.2 µg/g (46% AA reduction) to 0.46 µg/g (79% AA reduction), respectively (Table S3).

The acrylamide reduction potential of SlpA was also tested in the bread sample (Fig. 9b). The control sample reported an acrylamide content of 1.890 µg/g. Pre-treatment of dough with 0.5 U/g flour of SlpA did not result in a significant reduction of acrylamide content (7% AA reduction). However, a significant reduction of acrylamide content from 17 to 63% (p value < 0.05) was noted at SlpA concentrations ranging from 1 to 100 U/g (Table S3).

Discussion

Though L-asparaginases hold immense potential for the treatment of acute lymphoblastic leukemia (ALL), their use is often limited by the side effects that arise out of the intrinsic glutaminase activity of most asparaginases (Brumano et al. 2019; Castro et al. 2021). Consequently, a high affinity for L-asparagine (L-asparaginase activity) and the absence of L-glutaminase activity are important for the application of L-ASNase in cancer treatment. This manuscript describes the cloning and expression of an L-asparaginase contig, SlcA mined from whole metagenome shotgun sequencing data of *Sardinella longiceps* gut microbiota, in the quest for improved asparaginases for cancer therapy. Further, the study also evaluates the acrylamide mitigation potential of the expressed asparaginase, SlpA.

While the nucleic acid sequence of SlcA showed similarity to *Vibrio metoecus*, the deduced amino acid sequence showed similarity to L-ASNase from *Photobacterium leiognathi*. Phylogenetic analysis revealed the relatedness of SlpA to L-ASNases from *Vibrio campbellii* and *Photobacterium* sp. J15, belonging to phylum *Gammaproteobacteria*. *Photobacterium* sp., belonging to *Vibrionaceae* family, comprise both luminous and non-luminous

strains of Gram-negative bacteria that exist in symbiotic relationships with marine organisms (Urbanczyk et al. 2011). They have been reported to produce esterases, cold-active lipases, asparaginases, antimicrobial compounds, and polyunsaturated fatty acids (Moi et al. 2017). Production of L-ASNase by marine luminous bacteria was reported to be higher compared to that of asparaginases from other bacterial species such as *E. coli*, *Citrobacter freundii*, *Pseudomonas* sp., *Micrococcus* sp., *Bacillus* sp., and *Flavobacterium* sp. (Ramaiah and Chandramohan 1992). Till date, there is only one preliminary study on the expression of L-asparaginase gene of *Photobacterium* sp. strain J15. However, the study did not explore any application prospects of the expressed asparaginase (Yaacob et al. 2014).

Crude SlpA purified using Ni-NTA spin column did not exhibit asparaginase activity. The use of imidazole has been reported to result in the loss of enzyme activity (Gort and Maloy 1998). Even though imidazole is removed by dialysis, protein precipitation that occurs during the removal of > 100 mM imidazole is likely to render the protein inactive (Blowers 2000). SlpA was therefore purified to homogeneity using ammonium sulphate precipitation, dialysis, gel filtration, and ion exchange chromatography. Purified SlpA was observed to be a homodimer (74 kDa) upon polyacrylamide gel electrophoresis. Further, MALDI-TOF spectrometry and MASCOT analysis corroborated the monomer weight of SlpA revealed by SDS-PAGE analysis. Most bacterial asparaginases are homotetramers, though there are reports of monomeric, dimeric, and hexameric forms (Batool et al. 2016). Most type II L-ASNases from mesophilic sources exist as tetramers, with molecular weights ranging from 140 to 160 kDa, whereas type I L-ASNases are reportedly homodimers with molecular weights around 70 kDa (Izadpanah Qeshmi et al. 2018). L-asparaginase gene cloned and expressed from *Staphylococcus* sp. OJ82 in *E. coli* BL21 (DE3) presented a molecular mass of 37.5 kDa on SDS-PAGE analysis. However, gel filtration chromatography revealed the protein as a homodimer (Han et al. 2014). L-ASNase from *Enterobacter cloacae* was also reported to be a homodimer of 106 kDa molecular weight, with a subunit size of 52 kDa (Husain et al. 2016a).

Purified SlpA did not show any glutaminase activity. Intrinsic glutaminase activity of commercial L-ASNases contributes to side effects such as hypersensitive reactions and organ failure. Protein engineering has been proposed to overcome the limitations of current ASNase formulations. This is, however, a formidable task (Brumano et al. 2019). As L-asparaginase SlpA does not have any associated L-glutaminase activity, it is a promising alternative to existing L-ASNase drugs. SlpA displayed maximal asparaginase activity at pH 8. Though SlpA showed good activity at physiological pH, there was a complete loss of activity below pH

6. The maximum activity of L-ASNases observed at alkaline pH is on account of the balance between L-aspartic acid and L-aspartate. At acidic pH, L-aspartic acid has a high affinity for the active site of L-asparaginase, and hence, it becomes a competitive inhibitor (Miller et al. 1993). Under conditions of alkaline pH, the balance shifts towards L-aspartate which has less affinity to the active site, thereby promoting a favorable balance for the enzyme–substrate reaction (Lubkowski et al. 1994). The activity at physiological pH is a pre-requisite for the effective anticancer application of L-asparaginase. Most marine bacterial L-ASNases display optimal activity in the pH range of 6 to 9 (Izadpanah Qeshmi et al. 2018). L-asparaginases from *E. coli* and *E. chrysanthemi*, currently used for the treatment of acute lymphoblastic leukemia, reported optimal pH within the range of 6–8 and 8–9, respectively (Whelan and Wriston 1969; Capizzi et al. 1971; Nguyen et al. 2016). This implies that SlpA is an attractive candidate for tumor therapy.

SlpA showed maximal activity at 30 °C, in close agreement with the temperature optima of other bacterial asparaginases. For instance, type II L-ASNase from *E. coli* (ECII), currently used in cancer treatment, has a temperature optimum of 37 °C (Capizzi et al. 1971; Kumar and Sobha 2012). On the other hand, L-asparaginase from *E. chrysanthemi* (ErA) showed maximal activity at 50 °C (Kumar and Sobha 2012). Recombinant L-ASNase, produced by cloning and expression of L-asparaginase gene from *Pseudomonas fluorescens* in *E. coli* BL21, had a temperature optimum of 34 °C (Kishore et al. 2015). L-asparaginase produced by the marine isolate, *Paenibacillus barengoltzii* CAU904, showed a slightly higher temperature optimum of 45 °C (Shi et al. 2017). Though there are reports of heterologous expression of thermostable L-ASNases from *Thermococcus kodakarensis* KOD1 (Hong et al. 2014), *Pyrobaculum calidifontis* (Chohan et al. 2019), and *Palaeococcus ferrophilus* (Wang et al. 2021a) with temperature optima above 90 °C, they do not find much potential in cancer treatment. SlpA showed good affinity to L-asparagine with K_m and V_{max} values of 3.008 mM and 0.014 mM/min, respectively. This K_m value is less than the K_m values reported for ASNases from *E. coli* (3.5 mM) and *Erwinia carotovora* (7.14 mM) (Willis and Woolfolk 1974; Kamble et al. 2006), but higher than that of ASNases from *Vibrio succinogenes* and (0.0074 mM) and *B. licheniformis* (0.014 mM) (Mahajan et al. 2014). The low K_m value of SlpA shows better substrate affinity of SlpA compared to *E. coli* and *E. carotovora* asparaginases. Similarly, there are reports of asparaginases with slightly higher and lower K_m values compared to SlpA. L-asparaginase from *Halomonas elongata* showed K_m and V_{max} values of 5.6 mM and 2.2 $\mu\text{mol}/\text{min}$ (Ghasemi et al. 2017), while that from *Vibrio cholerae* recorded K_m and V_{max} values of 1.1 mM and 1006 $\mu\text{M}/\text{min}$, respectively (Radha et al. 2018).

The presence of L-asparagine is vital for normal functioning of human cells. In normal cells, the deficit of L-asparagine, induced by supplementation of L-asparaginase, can be circumvented by alternate pathways that synthesize L-asparagine from aspartate and glutamine using asparagine synthetase (Piatkowska-Jakubas et al. 2008). However, leukemia and lymphoma cells have low levels of asparagine synthetase expression and hence, depletion of L-asparagine from plasma leads to inhibition of DNA, RNA, and protein synthesis and causes apoptotic death (Fung and Chan 2017). The anticancer activity of L-ASNase also covers a wide variety of other cancer types such as breast cancer, glioma, ovarian cancer, prostate cancer, hepatocarcinoma, sarcoma, myeloma, and others (Covini et al. 2011; Abakumova et al. 2012; Chiu et al. 2020).

The cytotoxic activity of SlpA was evaluated against both leukemia and breast cancer cell lines. SlpA showed good cytotoxic activity against K-562 and MCF-7 cell lines with IC_{50} values of 0.3443 and 2.692 U/mL, respectively. L-ASNase purified from *Bacillus licheniformis* recorded IC_{50} values of 0.22 IU, 0.153 IU, and 0.78 IU for the leukemic cell lines, Jurkat clone E6-1, K562, and breast cancer cell line MCF-7, respectively (Mahajan et al. 2014). Recombinant *Rhodospirillum rubrum* L-asparaginase (RrA) has also been reported to show a significant cytotoxic effect against K-562, DU-145 (prostate cancer), MDA-MB-231 (breast cancer), and MCF-7 cell lines with IC_{50} values of 1.80 IU/mL, 9.19 IU/mL, 34.62 IU/mL, and 43.3 IU/mL, respectively (Pokrovskaya et al. 2012a, b). The lower IC_{50} values obtained for leukemic cell lines in comparison to other cell lines indicate a high effectiveness of asparaginase for the treatment of leukemia compared to other cancers. The cytotoxicity induced by SlpA was also confirmed by acridine orange/ethidium bromide (AO/EB) staining and SEM analysis of the treated and non-treated cells.

About 15–20% of patients treated with L-ASNase from *E. coli* reportedly develop hypersensitivity and toxicity to the drug (Mahajan et al. 2014). This reflects a cytotoxic effect on normal tissues. Hence, it is imperative to test the difference in sensitivity of a drug between normal and cancer cells. The IC_{50} concentrations of SlpA did not show any significant loss of viability of normal lymphocytes. Hemolysis is one of the main side effects of drugs used to treat hematologic malignancies (Lubran 1989). Both qualitative and quantitative hemolysis assays confirmed that SlpA is hemocompatible up to 40 U/mL concentration. While purified L-ASNase from *Bacillus tequilensis* PV9W showed hemocompatibility up to 6 IU (Shakambari et al. 2016), L-ASNases from *Enterobacter cloacae* and *Pseudomonas otitidis* were reported to be hemocompatible up to a concentration of 15 IU/mL (Husain et al. 2016a, b). Therefore, SlpA may be considered safe for in vivo application.

Pre-treatment of biscuit and bread dough with different concentrations of SlpA resulted in a clear, dose-dependent

reduction of acrylamide formation incumbent during the baking process. A 79% acrylamide reduction was caused by 100 U/g flour of SlpA in biscuit samples. Different dosages of L-ASNases have been used for acrylamide mitigation in biscuit samples. Semi-sweet biscuits prepared from flour treated with 0.525 and 1.05 ASNU/g of Acrylaway® L-asparaginase (from *Aspergillus oryzae*) resulted in 65% and 84% reduction in acrylamide contents, respectively. The use of ASNase dose of 1 ASNU/g flour resulted in a 90% reduction of acrylamide content in ginger biscuits. One ASNU is the amount of enzyme that produces 1 µmol of ammonia per minute at 37 °C and pH 7.0 using Acrylaway® (Novozymes A/S, Denmark) (Hendriksen et al. 2009). Short dough biscuits prepared with 0.5 U/g flour of Acrylaway® L-asparaginase caused acrylamide reduction ranging from 27 to 70% (Anese et al. 2011). In the present study, incorporation of 0.5 U/g flour of SlpA caused a 46% reduction in acrylamide formation. Addition of 0.5 U/g flour of L-ASNase from *Rhizomucor miehei* (RmAsnase) resulted in a 15% reduction of acrylamide. An 80% reduction of acrylamide was achieved with 10 U/g flour of RmAsnase (Huang et al. 2014).

L-asparaginase SlpA showed a lower acrylamide reduction potential in bread samples compared to biscuit samples. While the acrylamide degradation ratio in the bread sample was less than 30% up to 10 U/g flour concentrations of SlpA, 0.5 U/g flour concentration of SlpA resulted in a 46% reduction of acrylamide formation in the biscuit sample. The differential acrylamide mitigation potential of SlpA in bread and biscuit may be attributed to the difference in the dough pH of the samples. The pH of dough in the case of biscuit and bread samples were 8.6 and 5.9, respectively. The pronounced acrylamide mitigation observed in biscuit samples may therefore be attributed to the alkaline pH optima of SlpA (pH 8). Varying concentrations of L-ASNase from *Cladosporium* sp., ranging from 0.5 to 3 U/g flour, were used to reduce acrylamide formation in sweet bread. Addition of 3 U/g flour of L-asparaginase reduced acrylamide formation to 97% and 73% in the crust and crumb regions (Mohan Kumar et al. 2014). Bread dough treated with 0.5 and 10 U/g flour of RmAsnase resulted in the reduction of acrylamide concentrations to 60% and 80%, respectively (Huang et al. 2014).

From the results, it follows that glutaminase-free SlpA without cytotoxic effects on normal cells is an excellent candidate for development as a chemotherapeutic drug as well as a food processing aid for the baking industry. From a future perspective, the contig sequence SlcA may also be subjected to rational protein engineering for enhanced stability and substrate affinity of the encoding enzyme. Further, a successful acrylamide mitigation strategy using SlpA can be formulated after evaluation of the cumulative effects of the baking ingredients and conditions on the nutritional quality

of the end products. Additional research is underway to address these possibilities for the improvement of SlpA for its successful application in the pharmaceutical and baking industries.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-022-11954-7>.

Acknowledgements The authors gratefully acknowledge the project grants from Centre for Marine Living Resources and Ecology-Ministry of Earth Sciences, Government of India (No. MOES/10-MLR/2/2007 and MOES/10-MLRTD/03/2013) given to Dr. Sarita G Bhat. Author Tina Kollannoor Johny acknowledges UGC, Government of India for research fellowship.

Author contribution TKJ performed cloning, protein purification, and baking experiments and drafted the manuscript. RMP assisted in data analysis and manuscript preparation. BES and RSS assisted in protein purification experiments. UBS and SPP performed cell line experiments. SGB acquired funding and reviewed the manuscript. All the authors read and approved the final manuscript.

Data Availability The datasets generated or analyzed during the current study are available in the NCBI BioSample Submission Portal as BioProject PRJNA433183 and SRA accession number SRR6677342 and in NCBI GenBank under accession number MT740290.

Declarations

Ethics approval Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Abakumova OYu, Podobed OV, Karalkin PA, Kondakova LI, Sokolov NN (2012) Antitumor activity of L-asparaginase from *Erwinia carotovora* against different human and animal leukemic and solid tumor cell lines. *Biochem Moscow Suppl Ser B* 6:307–316. <https://doi.org/10.1134/S1990750812040026>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Anese M, Quarta B, Frias J (2011) Modelling the effect of asparaginase in reducing acrylamide formation in biscuits. *Food Chem* 126:435–440. <https://doi.org/10.1016/j.foodchem.2010.11.007>
- Batool T, Makky EA, Jalal M, Yusoff MM (2016) A comprehensive review on L-asparaginase and its applications. *Appl Biochem Biotechnol* 178:900–923. <https://doi.org/10.1007/s12010-015-1917-3>
- Blowers DP (2000) Affinity separation. In: Wilson ID (ed) *Encyclopedia of separation science*, 1st edn. Academic Press, pp 277–283
- Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J (2012) Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biol* 13:R122. <https://doi.org/10.1186/gb-2012-13-12-r122>
- Broome JD (1963) Evidence that the L-asparaginase of guinea pig serum is responsible for its antilymphoma effects. *J Exp Med* 118:99–120

- Brumano LP, da Silva FVS, Costa-Silva TA, Apolinário AC, Santos JHPM, Kleingesinds EK, Monteiro G, de Rangel-Yagui CO, Benyahia B, Junior AP (2019) Development of L-asparaginase biobetters: current research status and review of the desirable quality profiles. *Front Bioeng Biotechnol*. <https://doi.org/10.3389/fbioe.2018.00212>
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
- Cachumba JJM, Antunes FAF, Peres GFD, Brumano LP, Santos JCD, Da Silva SS (2016) Current applications and different approaches for microbial L-asparaginase production. *Braz J Microbiol* 47:77–85. <https://doi.org/10.1016/j.bjm.2016.10.004>
- Capizzi RL, Bertino JR, Skeel RT, Creasey WA, Zanes R, Olayon C, Peterson RG, Handschumacher RE (1971) L-asparaginase: clinical, biochemical, pharmacological, and immunological studies. *Ann Intern Med* 74:893–901. <https://doi.org/10.7326/0003-4819-74-6-893>
- Castro D, Marques AS, Almeida MR, de Paiva GB, Bento H, Pedrolli DB, Freire MG, Tavares AP, Santos-Ebinuma VC (2021) L-asparaginase production review: bioprocess design and biochemical characteristics. *Appl Microbiol Biotechnol* 105:4515–4534. <https://doi.org/10.1007/s00253-021-11359-y>
- Chiu M, Taurino G, Bianchi MG, Kilberg MS, Bussolati O (2020) Asparagine synthetase in cancer: beyond acute lymphoblastic leukemia. *Front Oncol*. <https://doi.org/10.3389/fonc.2019.01480>
- Chohan SM, Rashid N, Sajed M, Imanaka T (2019) Pcal_0970: an extremely thermostable L-asparaginase from *Pyrobaculum calidifontis* with no detectable glutaminase activity. *Folia Microbiol (praha)* 64:313–320. <https://doi.org/10.1007/s12223-018-0656-6>
- Covini D, Tardito S, Bussolati O, Chiarelli LR, Pasquetto MV, Digilio R, Scotti GVC (2011) Expanding targets for a metabolic therapy of cancer: L-asparaginase. In: *Recent Patents on Anti-Cancer Drug Discovery*. <http://www.eurekaselect.com/89172/article>. Accessed 14 Feb 2020
- Fung MKL, Chan GC-F (2017) Drug-induced amino acid deprivation as strategy for cancer therapy. *J Hematol Oncol* 10:144. <https://doi.org/10.1186/s13045-017-0509-9>
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784–3788
- Ghasemi A, Asad S, Kabiri M, Dabirmanesh B (2017) Cloning and characterization of *Halomonas elongata* L-asparaginase, a promising chemotherapeutic agent. *Appl Microbiol Biotechnol* 101:7227–7238. <https://doi.org/10.1007/s00253-017-8456-5>
- Gort S, Maloy S (1998) Purification of a hexahistidine-tagged protein using L-histidine as the eluent. *Tech Tips Online* 3:54–55. [https://doi.org/10.1016/S1366-2120\(08\)70098-9](https://doi.org/10.1016/S1366-2120(08)70098-9)
- Han S, Jung J, Park W (2014) Biochemical characterization of L-asparaginase in NaCl-tolerant *Staphylococcus* sp. OJ82 isolated from fermented seafood. *J Microbiol Biotechnol* 24:1096–1104. <https://doi.org/10.4014/jmb.1405.05021>
- Hendriksen HV, Kornbrust BA, Østergaard PR, Stringer MA (2009) Evaluating the potential for enzymatic acrylamide mitigation in a range of food products using an asparaginase from *Aspergillus oryzae*. *J Agric Food Chem* 57:4168–4176. <https://doi.org/10.1021/jf900174q>
- Hogervorst JG, Schouten LJ, Konings EJ, Goldbohm RA, van den Brandt PA (2007) A prospective study of dietary acrylamide intake and the risk of endometrial, ovarian, and breast cancer. *Cancer Epidemiol Biomarkers Prev* 16:2304–2313. <https://doi.org/10.1158/1055-9965.EPI-07-0581>
- Hong SJ, Lee YH, Khan AR, Ullah I, Lee C, Park CK, Shin JH (2014) Cloning, expression, and characterization of thermophilic L-asparaginase from *Thermococcus kodakarensis* KOD1. *J Basic Microbiol* 54:500–508. <https://doi.org/10.1002/jbom.201300741>
- Huang L, Liu Y, Sun Y, Yan Q, Jiang Z (2014) Biochemical characterization of a novel L-asparaginase with low glutaminase activity from *Rhizomucor miehei* and its application in food safety and leukemia treatment. *Appl Environ Microbiol* 80:1561–1569. <https://doi.org/10.1128/AEM.03523-13>
- Husain I, Sharma A, Kumar S, Malik F (2016a) Purification and characterization of glutaminase free asparaginase from *Enterobacter cloacae*: *in vitro* evaluation of cytotoxic potential against human myeloid leukemia HL-60 Cells. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0148877>
- Husain I, Sharma A, Kumar S, Malik F (2016b) Purification and characterization of glutaminase free asparaginase from *Pseudomonas otitidis*: Induce apoptosis in human leukemia MOLT-4 cells. *Biochimie* 121:38–51. <https://doi.org/10.1016/j.biochi.2015.11.012>
- IARC Summary Evaluation (1994) Acrylamide, vol 60. <http://www.inchem.org/documents/iarc/vol60/m60-11.html>. Accessed 24 April 2021
- Imada A, Igarasi S, Nakahama K, Isono M (1973) Asparaginase and glutaminase activities of micro-organisms. *Microbiology* 76:85–99. <https://doi.org/10.1099/00221287-76-1-85>
- Izadpanah Qeshmi F, Homaei A, Fernandes P, Javadpour S (2018) Marine microbial L-asparaginase: Biochemistry, molecular approaches and applications in tumor therapy and in food industry. *Microbiol Res* 208:99–112. <https://doi.org/10.1016/j.micres.2018.01.011>
- Jia R, Wan X, Geng X, Xue D, Xie Z, Chen C (2021) Microbial L-asparaginase for application in acrylamide mitigation from food: current research status and future perspectives. *Microorganisms* 9:1659. <https://doi.org/10.3390/microorganisms9081659>
- Jiang Z, Li X, Yang S, Li L, Tan S (2005) Improvement of the bread-making quality of wheat flour by the hyperthermophilic xylanase B from *Thermotoga maritima*. *Food Res Int* 38:37–43. <https://doi.org/10.1016/j.foodres.2004.07.007>
- Kamble VP, Rao RS, Borkar PS, Khobragade CN, Dawane BS (2006) Purification of L-asparaginase from a bacteria *Erwinia carotovora* and effect of a dihydropyrimidine derivative on some of its kinetic parameters. *Indian J Biochem Biophys* 43:391–394
- Kasibhatla S, Amarante-Mendes GP, Finucane D, Brunner T, Bossy-Wetzel E, Green DR (2006) Acridine orange/ethidium bromide (AO/EB) staining to detect apoptosis. *Cold Spring Harb Protoc* 2006:pdb.prot4493. <https://doi.org/10.1101/pdb.prot4493>
- Kidd JG (1953) Regression of transplanted lymphomas induced *in vivo* by means of normal guinea pig serum. *J Exp Med* 98:565–582
- Kishore V, Nishita KP, Manonmani HK (2015) Cloning, expression and characterization of L-asparaginase from *Pseudomonas fluorescens* for large scale production in *E. coli* BL21. *3 Biotech* 5:975–981. <https://doi.org/10.1007/s13205-015-0300-y>
- Kumar DS, Sobha K (2012) L-Asparaginase from microbes: a comprehensive review. *Adv Biore* 3:137–157
- Kumar J, Das S, Teoh SL (2018a) Dietary acrylamide and the risks of developing cancer: facts to ponder. *Front Nutr*. <https://doi.org/10.3389/fnut.2018.00014>
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018b) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Lineback DR, Coughlin JR, Stadler RH (2012) Acrylamide in foods: a review of the science and future considerations. *Annu Rev Food Sci Technol* 3:15–35. <https://doi.org/10.1146/annurev-food-022811-101114>
- Lubkowski J, Wlodawer A, Housset D, Weber IT, Ammon HL, Murphy KC, Swain AL (1994) Refined crystal structure of *Acinetobacter glutaminasificans* glutaminase-asparaginase. *Acta Crystallogr D* 50:826–832. <https://doi.org/10.1107/S0907444994003446>
- Lubran MM (1989) Hematologic side effects of drugs. *Ann Clin Lab Sci* 19:114–121
- Mahajan RV, Kumar V, Rajendran V, Saran S, Ghosh PC, Saxena K (2014) Purification and characterization of a novel and robust L-asparaginase having low-glutaminase activity from *Bacillus*

- licheniformis*: *in vitro* evaluation of anti-cancerous properties. PLoS ONE 9:9. <https://doi.org/10.1371/journal.pone.0099037>
- McGahon AJ, Martin SJ, Bissonnette RP, Mahboubi A, Shi Y, Mogil RJ, Nishioka WK, Green DR (1995) Chapter 9 the end of the (cell) line: methods for the study of apoptosis *in vitro*. In: Schwartz LM, Osborne BA (eds) *Methods in Cell Biology*. Academic Press, pp 153–185
- Medeiros Vinci R, Mestdagh F, De Meulenaer B (2012) Acrylamide formation in fried potato products – present and future, a critical review on mitigation strategies. *Food Chem* 133:1138–1154. <https://doi.org/10.1016/j.foodchem.2011.08.001>
- Miller M, Rao JK, Wlodawer A, Gribskov MR (1993) A left-handed crossover involved in amidohydrolase catalysis. Crystal structure of *Erwinia chrysanthemi* L-asparaginase with bound L-aspartate. *FEBS Lett* 328:275–279. [https://doi.org/10.1016/0014-5793\(93\)80943-o](https://doi.org/10.1016/0014-5793(93)80943-o)
- Mohan Kumar NS, Shimray CA, Indrani D, Manonmani HK (2014) Reduction of acrylamide formation in sweet bread with L-asparaginase treatment. *Food Bioprocess Technol* 7:741–748. <https://doi.org/10.1007/s11947-013-1108-6>
- Moi IM, Roslan NN, Leow ATC, Ali MSM, Rahman RNZR Abd, Rahimpour A, Sabri S (2017) The biology and the importance of *Photobacterium* species. *Appl Microbiol Biotechnol* 101:4371–4385. <https://doi.org/10.1007/s00253-017-8300-y>
- Nguyen TTH, Nguyen CT, Nguyen TSL, Do TT (2016) Optimization, purification and characterization of recombinant L-asparaginase II in *Escherichia coli*. *Afr J Biotechnol* 15:1681–1691. <https://doi.org/10.5897/AJB2016.15425>
- Nguyen HT, van der Fels-Klerx HJI, van Boekel MAJS (2017) Acrylamide and 5-hydroxymethylfurfural formation during biscuit baking. Part II: Effect of the ratio of reducing sugars and asparagine. *Food Chem* 230:14–23. <https://doi.org/10.1016/j.foodchem.2017.03.009>
- Nguyen HA, Su Y, Zhang JY, Antanasijevic A, Caffrey M, Schalk AM, Liu L, Rondelli D, Oh A, Mahmud DL, Bosland MC, Kajdacsy-Balla A, Peirs S, Lammens T, Mondelaers V, De Moerloose B, Goossens S, Schlicht MJ, Kabirov KK, Lyubimov AV, Merrill BJ, Saunthararajah Y, Van Vlierbergh P, Lavie A (2018) A novel L-asparaginase with low L-glutaminase coactivity is highly efficacious against both T- and B-cell acute lymphoblastic leukemias *in vivo*. *Cancer Res* 78:1549–1560. <https://doi.org/10.1158/0008-5472.CAN-17-2106>
- Noguchi H, Taniguchi T, Itoh T (2008) MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* 15:387–396. <https://doi.org/10.1093/dnares/dsn027>
- Piatkowska-Jakubas B, Krawczyk-Kuliś M, Giebel S, Adamczyk-Cioch M, Czyz A, Lech Marañda E, Paluszewska M, Pałynyczko G, Piszcz J, Hołowicki J, Polish Adult Leukemia Group (2008) Use of L-asparaginase in acute lymphoblastic leukemia: recommendations of the Polish adult leukemia group. *Pol Arch Med Wewn* 118:664–669
- Pokrovskaya MV, Aleksandrova SS, Pokrovsky VS, Omeljanjuk NM, Borisova AA, Anisimova NY, Sokolov NN (2012a) Cloning, expression and characterization of the recombinant *Yersinia pseudotuberculosis* L-asparaginase. *Protein Expr Purif* 82:150–154. <https://doi.org/10.1016/j.pep.2011.12.005>
- Pokrovskaya MV, Pokrovskiy VS, Aleksandrova SS, Anisimova NYu, Andrianov RM, Treschalina EM, Ponomarev GV, Sokolov NN (2012b) Recombinant intracellular *Rhodospirillum rubrum* L-asparaginase with low L-glutaminase activity and antiproliferative effect. *Biochem Moscow Suppl Ser B* 6:123–131. <https://doi.org/10.1134/S1990750812020096>
- Radha R, Arumugam N, Gummedi SN (2018) Glutaminase free L-asparaginase from *Vibrio cholerae*: heterologous expression, purification and biochemical characterization. *Int J Biol Macromol* 111:129–138. <https://doi.org/10.1016/j.ijbiomac.2017.12.165>
- Ramaiah N, Chandramohan D (1992) Production of L-asparaginase by the marine luminous bacteria. *Indian J Mar Sci* 21:212–214
- Ramya LN, Doble M, Rekha VPB, Pulicherla KK (2012) L-asparaginase as potent anti-leukemic agent and its significance of having reduced glutaminase side activity for better treatment of acute lymphoblastic leukaemia. *Appl Biochem Biotechnol* 167:2144–2159. <https://doi.org/10.1007/s12010-012-9755-z>
- Rosell CM, Collar C (2009) Effect of temperature and consistency on wheat dough performance. *Int J Food Sci* 44:493–502. <https://doi.org/10.1111/j.1365-2621.2008.01758.x>
- Shakambari G, Birendranarayan AK, Lincy MJA, Rai SK, Ahamed QT, Ashokkumar B, Saravanan M, Mahesh A, Varalakshmi P (2016) Hemocompatible glutaminase free L-asparaginase from marine *Bacillus tequilensis* PV9W with anticancer potential modulating p53 expression. *RSC Adv* 6:25943–25951. <https://doi.org/10.1039/C6RA00727A>
- Shi R, Liu Y, Mu Q, Jiang Z, Yang S (2017) Biochemical characterization of a novel L-asparaginase from *Paenibacillus barengoltzii* being suitable for acrylamide reduction in potato chips and mooncakes. *Int J Biol Macromol* 96:93–99. <https://doi.org/10.1016/j.ijbiomac.2016.11.115>
- Stadler RH, Robert F, Riediker S, Varga N, Davidek T, Devaud S, Goldmann T, Hau J, Blank I (2004) In-depth mechanistic study on the formation of acrylamide and other vinyllogous compounds by the maillard reaction. *J Agric Food Chem* 52:5550–5558. <https://doi.org/10.1021/jf0495486>
- Urbanczyk H, Ast JC, Dunlap PV (2011) Phylogeny, genomics, and symbiosis of *Photobacterium*. *FEMS Microbiol Rev* 35:324–342. <https://doi.org/10.1111/j.1574-6976.2010.00250.x>
- Van Trimpont M, Peeters E, De Visser Y, Schalk AM, Mondelaers V, De Moerloose B, Lavie A, Lammens T, Goossens S, Van Vlierbergh P (2022) Novel insights on the use of L-asparaginase as an efficient and safe anti-cancer therapy. *Cancers* 14:902. <https://doi.org/10.3390/cancers14040902>
- Wang Y, Wu H, Zhang W, Xu W, Mu W (2021a) Efficient control of acrylamide in French fries by an extraordinarily active and thermo-stable L-asparaginase: a lab-scale study. *Food Chem* 360:130046. <https://doi.org/10.1016/j.foodchem.2021.130046>
- Wang Y, Xu W, Wu H, Zhang W, Guang C, Mu W (2021b) Microbial production, molecular modification, and practical application of L-asparaginase: a review. *Int J Biol Macromol* 186:975–983. <https://doi.org/10.1016/j.ijbiomac.2021.07.107>
- Warangkar SC, Khobragade CN (2010) Purification, characterization, and effect of thiol compounds on activity of the *Erwinia carotovora* L-asparaginase. In: *Enzyme Research*. <https://www.hindawi.com/journals/er/2010/165878/>. Accessed 1 Feb 2020
- Whelan HA, Wriston JC (1969) Purification and properties of asparaginase from *Escherichia coli* B. *Biochemistry* 8:2386–2393. <https://doi.org/10.1021/bi00834a020>
- WHO model list of essential medicines- 21st List <https://www.who.int/publications-detail-redirect/WHOMVPPEMPIAU2019.06>. Accessed 24 Apr 2021
- Willis RC, Woolfolk CA (1974) Asparagine utilization in *Escherichia coli*. *J Bacteriol* 118:231–241
- Yaacob M, Hasan W, Ali M, Rahman R, Salleh A, Basri M, Leow T (2014) Characterisation and molecular dynamic simulations of J15 asparaginase from *Photobacterium* sp. strain J15. *Acta Biochim Pol* 61:745–752

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



An Improvised Random Forest Model for Breast Cancer Classification

Dr. Tina Elizabeth Mathew^{1*}

Abstract

Breast Cancer is considered as the most common cancer in females with high incidence rate. The evolution of modern facilities has helped in reducing the mortality rate, yet the incidence is still the highest among all cancers affecting women. Early diagnosis is a predominant factor for survival. Hence techniques to assist the current modalities are essential. Machine learning techniques have been used so as to produce better prediction and classification models which will aid in better and earlier disease diagnosis and classification. Random Forest is a supervised machine learning classifier that helps in better classification. Random Forests are applied to the Wisconsin breast cancer dataset and the performance of the classifier is evaluated for breast cancer classification. Here in this study an improvised random forest model which uses a cost sensitive learning approach for classification is proposed and it is found to have a better performance than the traditional random forest approach. The model gave an accuracy of 97.51%.

Key Words: Cost Matrix, Decision Trees, Breast Cancer, Classification, Improved Random Forest Classifier Approach (IRFC). 713

DOI Number: 10.14704/nq.2022.20.5.NQ22227

NeuroQuantology 2022; 20(5):713-722

Introduction

Breast Cancer, the neoplasm of the breast is a prevalent disease among women which is of great concern to women of all ages, countries and ethnicity. Recent statistics published by the World Health Organization in 2021 show that it is ranked first in incidence and first or second in mortality rate in almost all countries of the world. Besides the modern medical modalities available supplementary techniques can be used for early detection of the disease which is a key factor for survival. Several medical modalities like Mammograms, MRI, CT scan, Thermography, USG Scan and many hybrid state of art techniques are available for disease diagnosis, but each of these modalities have their own pros and cons. A major issue is that these techniques use detrimental x

rays and this is harmful for the patient. Besides the diagnosis process can be painful as well as stressful to the patients. The diagnostic accuracy can also at times be inconclusive and incorrect, Hence, to avoid these issues alternative techniques can be used to assist the medical practitioner. Literature studies indicate the use of various machine learning techniques for better identification, prediction and classification of the disease. (Algehyne et al, 2022), (Balaraman, 2020), (Mathew, 2019a), (Mathew, 2019b). Machine learning is a subdomain of artificial intelligence and to unravel complex patterns from the heterogenous biological data, machine learning methods can be employed [Qusit et al, 2021].

Corresponding author: Dr. Tina Elizabeth Mathew

Address: ^{1*}Assistant Professor in Computer Science, Government College Kariavattom, Thiruvananthapuram, Kerala, India.

^{1*}E-mail: tinamathew04@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2022 **Accepted:** 10 May 2022



Machine learning techniques can be classified into supervised and unsupervised techniques and each category has a plethora of methods. These techniques have seen to provide effective classification of diseases as per (Algehyne et al, 2022), (Khourdifi, Bahaj 2018), (Mathew, 2019c), (Mathew, Kumar, 2020), (Mathew, Kumar, 2021). Decision Trees are techniques that belong to the category of supervised techniques. Decision Trees are considered to be simple methods yet powerful enough for classification. Two broad categories of Decision Trees are classification trees and regression trees. Classification trees are further subdivided into various types such as ID3, C4.5, CART, CHAID, MARS and so forth. A major drawback with decision trees is that they can overfit. So usually to solve this concern ensembles or groups of decision trees are taken as classifiers. Four such categories are Bagged trees, Random Forest, Boosted Trees and Rotation Forest. Each category exhibits suitability for classification (Mashudi et al, 2021).

Random Forest is considered to be robust than individual decision trees. Random forest takes a forest of trees and trains them using the bootstrap aggregation technique. The prediction of each tree is taken and the majority votes of all these predictions is produced as the final output. Bagging takes different subsets of samples instead of using one single set. The Random Forest classifiers are seen to provide a moderately high accuracy without the need of normalization of any dataset values (Ara et al, 2021). Albeit, an issue is that they are black box techniques and reasons behind the diagnosis or classification cannot be explained (wang et al, 2020).

Objectives of the Study

- To develop an improved Random Forest Model using cost sensitive learning.
- To compare the performance of the traditional Random Forest Model with the proposed improved model.
- To identify aptness of the proposed model for breast cancer classification.

Significance of Study

- The intent is to identify and develop a supervised machine learning model for breast cancer classification so that medical practitioners can use this model as a diagnostic assistive technology.

- These assistive models are non-invasive, painless and harmless techniques. By utilizing these for the detection of breast cancer the patients need not suffer the painstaking disease detection processes which make use of harmful radiation.
- The paper is organized as follows, Section 1 represents the introduction, Section 2 represents the literature survey, while Section 3 proposes the materials used and methods applied, Section 4 represents the results and discussions, and the final Section 5 proposes the conclusion followed by references.

Literature Review

Various literature reviewed based on random forest models and its application in the breast cancer domain as well as other fields are given in the following section. In their work (Alam et al, 2019) used a feature ranking and selection strategy along with Random Forest and it was seen to provide better performance. The model was tested with 10 datasets and the random forest model was seen to perform consistently across datasets when compared to other classifiers such as SVM and Bayes Networks. In his work (Al-Quraishi et al, 2018) proposed methods where Random Forest was utilized in conjunction with other techniques for breast cancer risk prediction, assessing recurrence probability and development of a prediction model for forecasting survivability status. (Ara et al, 2021) in their work developed an automatic breast cancer diagnostic system using SVM and Random Forest and the model was seen to produce an accuracy of 96.5%. (Buttan et al, 2021), proposed an RF model with grid search to predict breast cancer and analysed the outcomes. In their work (Chaudhary et al, 2016), used feature selection using three attribute evaluators. The classifier is improved using an attribute evaluator method and an instance filter method. For risk assessment of breast cancer (Housseinpour et al, 2022), used an improved random forest algorithm and obtained promising results using the model. (Jackins et al, 2021), compared performance of random forests, naïve bayes, k-means and DBSCAN on various datasets and they concluded that random forests worked well with all datasets. (Jadhav et al, 2019), illustrated that random forests performed better than logistic regression and decision tree models. In their work (Kaur et al, 2019), proposed an IOT based smart health system



with random forest classifier and the results of the study were seen promising with the classifier giving good accuracy for various disease datasets. (Keles, 2019),] in their comparative study on various classifiers illustrated that random forest could achieve an accuracy above 90% for disease classification. Work proposed by (Li, Chen, 2018), also indicated that random forests can be suitable for breast cancer classification. (Macaulay, 2021) proposed a risk prediction model for breast cancer in African women and their proposed model used random forest to identify risk factors and it was found suitable for identifying them. Combining feature selection with the models were seen to improve model performance. (Rohan, 2019), combined random forest with Adaboost and illustrated that this helped in performance enhancement of the model. (Shaik, Srinivasan, 2019), in their work highlighted the need for accuracy improvement using Random Forests. Their work was solely focused on random forests. (Shahhoseini, Hu, 2020), improved the random forest classifier performance by using a weighted model. In their work (Sharma et al, 2018), compared various machine learning techniques for Breast cancer classification. [Zhu et al, 2018] produced a forecasting model for breast cancer prediction using a Random Forest- Adaboost combination.

A major limitation identified in the various literature is that as the number of trees used in the forest grows the model gets slower. And this affects the performance in real time predictions. Even then it is considered to be much better than other classifiers, specifically other decision trees as it can handle large datasets and also takes care of overfitting which is a factor that usually affects the performance of models. It helps tackle the problem of variance.

A main challenge faced in classification is class imbalance. There is no unique judgement about the degree of the imbalance that exist between class cardinalities. Some researchers have studied datasets that have not so severe imbalance where one class is few times smaller than other class, while others have considered more severe imbalance ratios such as 1:100, 1:1000 and so forth. A class with few examples does not help to find the data regularities.

There are several techniques to deal with this. Two broad classifications are databased and algorithmic based sampling methods. Besides this, literature

also recommends feature selection, and ensembling of techniques to deal with data imbalance.

In databased techniques the imbalanced data ratio is reduced by either adding more minority instances, known as oversampling or discarding some of the majority instances called under sampling. (Yap, et al, 2014). These are usually applied at the data pre-processing phase. This does not affect the algorithms used. In algorithmic based approach algorithms are modified internally to achieve this (Napierała, 2012).

Materials and Methods

About the Dataset

The WBCD dataset used in this study taken from the University of Wisconsin Hospitals contains 699 instances, reported till July 15th, 1992. This dataset contains sample code number and 10 attributes for each instance of disease. The independent attributes of the dataset are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses. All of them are represented by values within the range of 1 and 10. The dependent attribute is class, which is represented by integer value 2 and 4, where 2 stands for benign tumours and 4 stands for malignant tumours.

Random Forest

Random Forest is a popular supervised machine learning algorithm that is used for several sorts of classification problems (Huljanah et al, 2019), (Islam et al, 20200), (Jayaraj, Sathiamoorthy, 2019). It is an ensemble of tree-structured classifiers. Each tree of the forest outputs a vote, and it assigns each instance of the input to the most probable class label. RF which basically uses the Classification and Regression Tree (CART) algorithm is seen to employ a number of decision trees as weak classifiers (Ghiasi, Zendehboudi, 2021)]. But in contrast with CART it does not employ a greedy algorithm (Kumar, Poonkudi, 2019). Many studies show that Random Forest is a fast method, robust to noise and it is an efficacious ensemble that identifies non-linear patterns in data. It can handle both numerical as well as categorical data easily. One of the major advantages of Random Forest is that it does not suffer from over fitting, even when more trees get appended to the forest. Random Forests are considered to be a powerful technique for classification, pattern recognition so on and so



forth (Fruend, Mason, 1999), (Ganggayah et al, 2019). Its parallel architecture makes it a fast classifier [31] and it takes care of data imbalance (Khoshgoftaar et al, 2007).

The Parameter used in random forests is the n_{trees} , number of trees that constitute the forest. Here it is taken as 100. Since a random forest creates ensembles of multiple decision trees, hyperparameters are used to control the number of trees (Gupta, Garg, 2020) and they are:

1. $max_features$ (number of attributes to be selected from data for randomisation)
2. max_depth (for pre-pruning of trees)
3. $max_features = \sqrt{n_features}$ (for classification)

Applying Ensemble learning techniques enhances the performance of predictive models by improving their accuracy (Yifan, 2021).

Random Forest builds a tree as an ensemble of decision trees that uses bagging together with random sampling of training points. It consists of an uncorrelated forest of trees and builds multiple decision trees and merges them which helps the classifier in getting a more accurate and stable prediction. It is capable of handling missing values but tends to show bias towards the majority class. Hence to improve the classification of the minority class improvement is to be effected. Besides, even though Random Forests methods can handle imbalanced data as the imbalance rate increases classification ability can decrease so methods to counteract this can be used (zhu et al, 2018). Hence to overcome this an improved random Forest approach is proposed in the next section.

Improved Random Forest Approach

Improved-RFC approach uses the Random Forest algorithm, and a cost matrix that penalizes each misclassification of the positive class or minority class twice thus to give a cost sensitive Random Forest model. That provides better classification for the minority class which is the positive class in this dataset.

Cost Sensitive classification using a cost matrix can be categorized as an algorithmic based approach. Misclassification cost is associated with classes. A cost matrix proposes a means to differentiate between and highlight the importance of the two classification error categories -Type I and Type II. The cost matrix used in the IRFDC approach is represented in Figure 1. The cost weights that lead to optimal performance in the classification process is learned by using grid search technique.

0	1
2	0

Figure 1. Cost matrix

The aim of the approach is to improve classification accuracy of the traditional Random Forest algorithm for binary classification and reduce misclassification of the minority class aka. positive class.

The pseudo code of the Improved Random Forest Classifier approach is given below. Figure 2 gives the working architecture of the proposed model.

Algorithm of IRFC
Input: $D_{Train} = \{x_1, x_2 \dots x_n\}$ // Training dataset which includes a set of training examples with their class labels.
Output: Classification + Performance measures
Method:
Step 1: Partition dataset into training and testing sets.
Step 2: Select Random Forest Classifier for classification.
Step 3: Train the classifier. Use 10-fold CV. Check Accuracy measure. Apply grid search to get the optimal value for the cost matrix
Step 4: Apply cost matrix for each misclassified positive sample.
Step 5: Apply classifier on test data
Step 6: Output Performance measures- Accuracy, MCC, F Score, Kappa Statistic

716

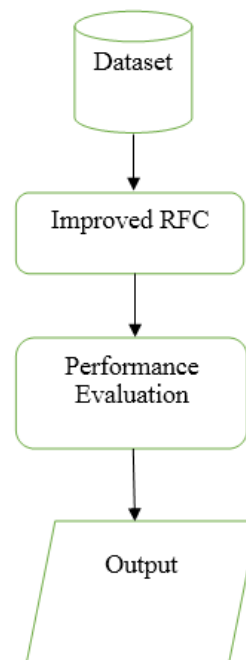


Figure 2. Architecture of IRFC Approach



Results & Discussion

The results obtained are shown in Table 1. The Improved Random Forest Approach enhanced accuracy to 97.51% from 96.63%. The misclassification of the positive class was much reduced from 12 instances to 5 instances.

The Matthews Correlation Coefficient (MCC) which is a measure of the quality of the binary classifier has shown a significant improvement. The IRFC model gave a value of 0.946 over 0.926. MCC is considered to be a more reliable measure than Accuracy.

Kappa Statistic of IRFC also displays a better value than that of the traditional approach. F measure also improved by a value of 0.009.

By applying the strategy of 10-fold cross validation overfitting is avoided.

P-R AUC and ROC AUC values of both models are also displayed.

Table 1. Performance Metrics of IRFC

Performance Metrics	Standard Random Forest	Improved Random Forest
Accuracy	96.6325	97.51
Kappa statistic	0.9259	0.9457
Confusion Matrix	433 11 12 227	432 12 5 234
F measure	0.966	0.975
MCC	0.926	0.946
ROC	0.992	0.992
PRC	0.990	0.991
Cost Matrix used		0 1
Time to build model(sec)	-0.04	2 0 0.03

A comparison of the proposed and traditional model in terms of Precision, Recall and Specificity is illustrated in Figure 3. For Precision and Recall the IRFC approach is superior in performance. While in Specificity the traditional model has a slightly higher value.

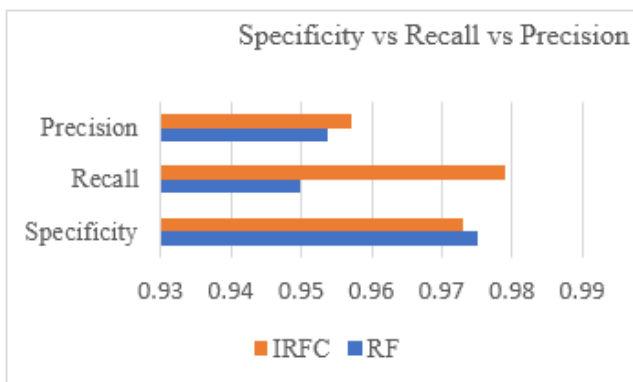


Figure 3. Specificity vs Recall vs Precision

Comparison of FPR and FNR of the models are illustrated in Figure 4. The false negative rate also termed as the miss rate. It denotes the probability that a true positive will be missed by the test. FNR of IRFC is superior than that of RF method. In case of FPR that of the RF method is slightly better than IRFC approach. An ideal model should have very low scores on FNR and FPR, while a practical model often has to make a trade-off between these two scores. The proposed IRFC has a better value for FNR.

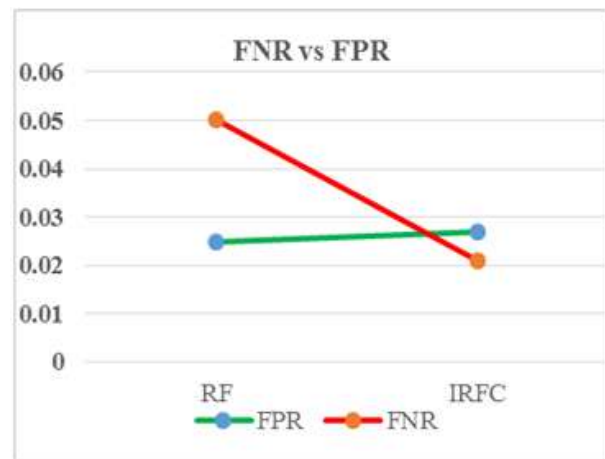


Figure 4. FNR vs FPR

The confusion matrix of the IRFC model depicts that the number of false negatives is reduced drastically, yet number of false positives is to be improved. The false negatives were much reduced from 12 instances to 5 instances. However, the false positives were not improved.

Besides this, the time taken to build the proposed model was seen to be lesser than the time taken by the standard model.

The experiment demonstrates that by integrating cost-sensitive learning to random forests effectively improves the classification performance.

The IRFC model is compared against other decision tree classifiers and figure 5 illustrates this comparison.



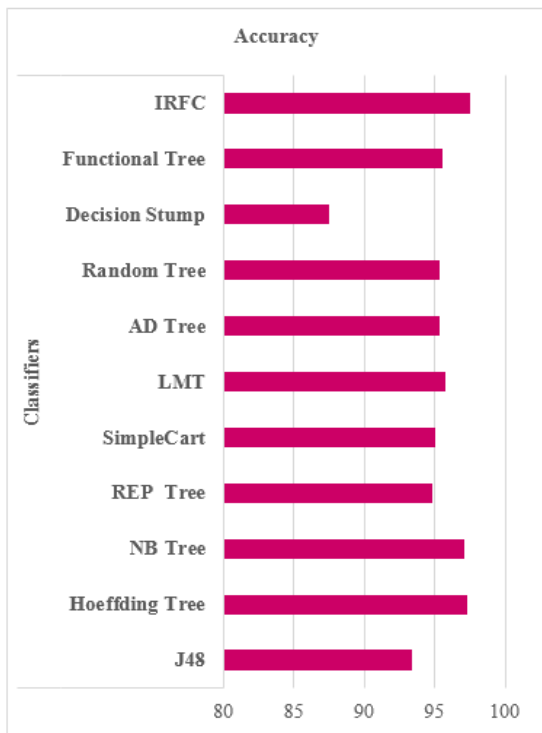


Figure 5. Comparison of Decision Trees

From the figure we can compare the accuracy measure obtained for various types of Decision Trees. The least accuracy is shown by the Decision Stump Model with a value of 87.5%. The proposed model has an accuracy of 97.51%. Other models like Hoeffding Tree and BB Tree show accuracy of

97.36% and 97.07% respectively. This highlights the superior performance of the proposed model amongst various decision Tree Classifiers. The proposed IRFC model is compared against few other classifiers- Support Vector Machines, k-Nearest Neighbours, Logistic Regression and figure 6 depicts the comparison done. The proposed method was seen to be superior in accuracy when compared with these classifiers.

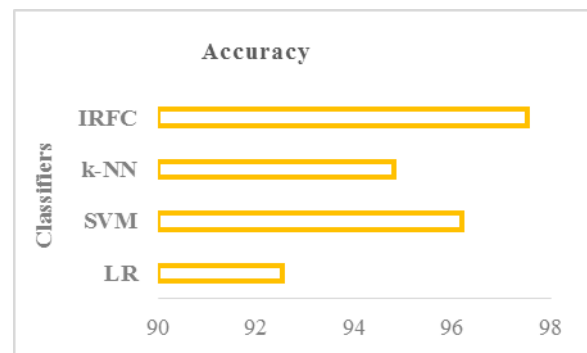


Figure 6. Comparison of Classifiers

The accuracy of the logistic Regression classifier was seen to be the lowest at 92.58%. The k-Nearest Neighbour classifier produced an accuracy of 94.8% and the Support Vector machine classifier gave an accuracy of 96.19%. The results highlight that the proposed IRFC model with accuracy of 97.515 was much superior in performance.

718

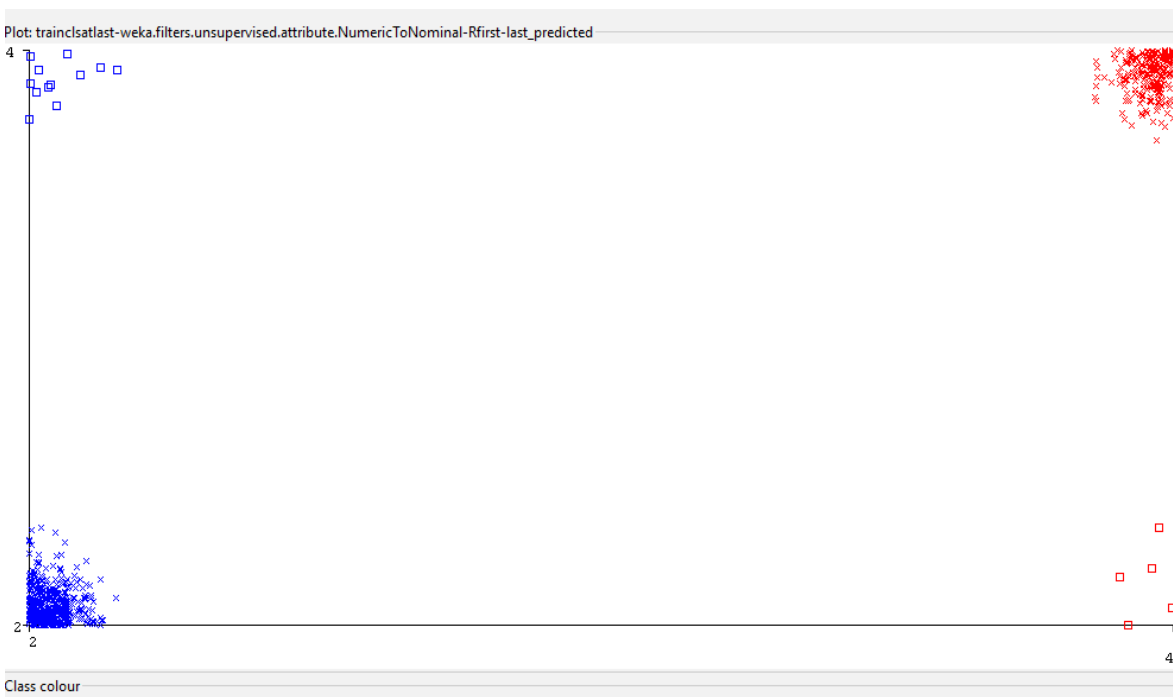


Figure 7. Visualization of Model Misclassification



Figure 7 gives the visualization of the classification errors made. Along the x axis the right most end shows the misclassified 5 positive instances. Similarly, along the y axis at the top end we can see the misclassified actual negative 12 instances. The P-R curve of the malignant class of the proposed Model is depicted in figure 8. The P R

curve uses recall on x-axis and precision on y axis. It is a helpful metric in cases where the data is imbalanced. Large P-R AUC values indicate better performance of the model. This is illustrated by the curve moving up towards the top right to left corner.

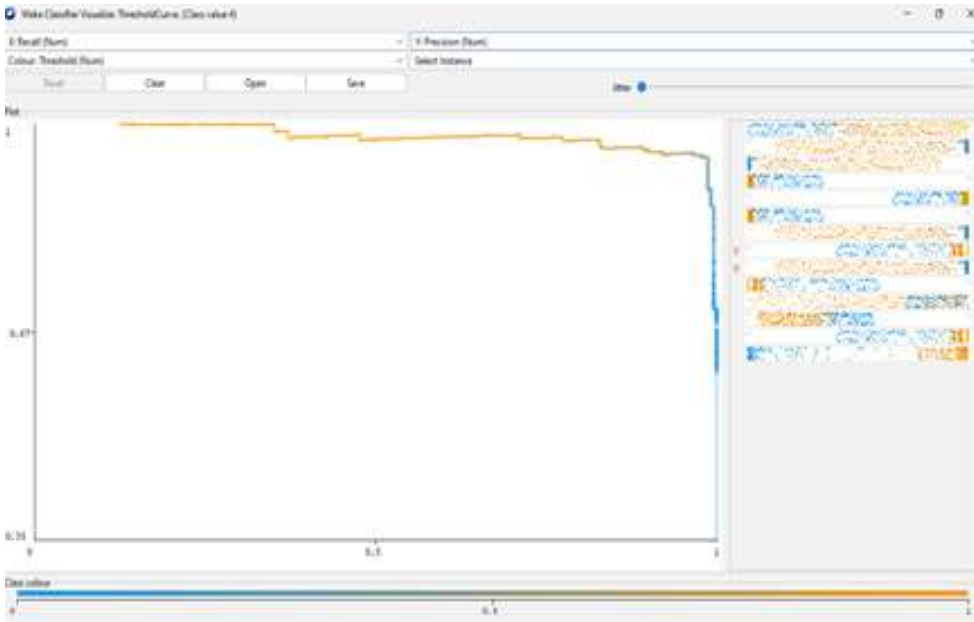


Figure 8. P-R Curve

Figure 9 depicts the ROC curve of the proposed IRFC model for the malignant class. ROC plots FPR on x axis and TPR on y axis. It is a useful metric to illustrate the diagnostic capability of the binary

classifier. An ideal ROC curve will be found along the upper left to right corner of the plot. The AUC ROC obtained is 0.9917

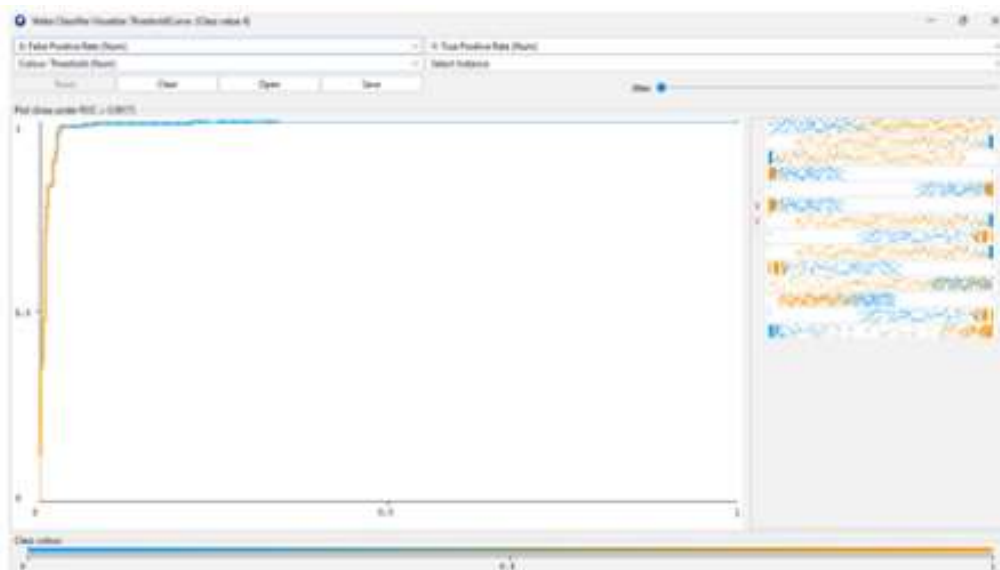


Figure 9. ROC Curve



The proposed model is compared with state of art technologies available in literature in Table 2. The proposed model is seen to outperform the models. (Imran et al, 2022) used an improved Random Forest Model and obtained an accuracy of 96%. (Zheng et al, 2020) used a hybrid Deep learning and Adaboost model and produced an accuracy of 97.2%. The proposed IRFC model produced an accuracy of 97.51%.

Thus, it can be comprehended that the proposed model is superior to these methods and can be used for breast cancer classification with a reduced misclassification of the positive class.

Table 2. Comparison with Literature

Author	Method	Accuracy
Imran et al, 2022	Random Forest	96
Zheng, 2020	Deep learning _ Adaboost	97.2
Proposed Model	IRFC	97.51

Conclusion and Outlook

Classification techniques have gained more traction with the availability of various kinds of clinical data, genomics data, omics data and many more. The paper discusses an approach for better classification accuracy on clinical data. The proposed IRFC approach improved the accuracy of classification over the traditional approach. The false negative rate was much reduced. However, the false positives produced by the approach is still a concern and other approaches are to be applied to provide better performance. In this context a suggestion will be to utilize optimization techniques, feature selection methods to improve the diagnostic accuracy. A problem with machine learning techniques is that they are data dependent and as data changes the structure and parameters required can change. To overcome this problem the models can be employed over different datasets of different data sizes so as to be updated accordingly.

Acknowledgements

I am grateful to Dr William H Wolberg of the University of Wisconsin Hospitals, Madison for the dataset provided.

Abbreviation List

- BC- Breast Cancer
- CV- Cross Validation
- FNR- False Negative Rate
- FPR- False Positive Rate
- IRFC- Improved Random Forest approach
- MCC- Matthews Correlation Coefficient
- P-R- Precision- Recall
- RF- Random Forest
- ROC- Receiver Operating Characteristics

References

Alam, M.Z., Rahman, M.S., and Rahman, M.S. 2019. A Random Forest based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*, 15, 100180.

Algehyne, E.A., Jibril, M.L., Algehainy, N.A., Alamri, O.A., and Alzahrani, A.K. 2022. Fuzzy Neural Network Expert System with an Improved Gini Index Random Forest-Based Feature Importance Measure Algorithm for Early Diagnosis of Breast Cancer in Saudi Arabia. *Big Data and Cognitive Computing*, 6(1), 13.

Al-Quraishi, T., Abawajy, J.H., Chowdhury, M. U., Rajasegarar, S., and Abdalrada, A.S. 2018, February. Breast cancer recurrence prediction using random forest model. *In International Conference on Soft Computing and Data Mining*, 318-329.

Ara, S., Das, A., and Dey, A. 2021. Malignant and benign breast cancer classification using machine learning algorithms. *In 2021 International Conference on Artificial Intelligence (ICAI)*, 97-101.

Balaraman, S. 2020. *Comparison of Classification Models for Breast Cancer Identification using Google Colab*. IEEE.

Buttan, Y., Chaudhary, A., and Saxena, K. 2021. An improved model for breast cancer classification using random forest with grid search method. *In Proceedings of Second International Conference on Smart Energy and Communication*, 407-415.

Chaudhary, A., Kolhe, S., and Kamal, R. 2016. An improved random forest classifier for multi-class classification. *Information Processing in Agriculture*, 3(4), 215-222.

Freund, Y. and Mason, L. 1999, The Alternating Decision Tree Learning Algorithms S.N. www1.Cs.Columbia.Edu/Compbio/Medusa/Non_Html_Files/Freund_Atrees.pdf

Ganggayah, M.D., Taib, N.A., Har, Y.C., Lio, P., and Dhillon, S.K. 2019. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC medical informatics and decision making*, 19(1), 1-17.

Ghiasi, M.M., and Zendehboudi, S. 2021. Application of decision tree-based ensemble learning in the classification of breast cancer. *Computers in Biology and Medicine*, 128, p 104089.

Gupta, P., and Garg, S. 2020. Breast cancer prediction using varying parameters of machine learning models. *Procedia Computer Science*, 171, 593-601.

Hosseinpour, M., Ghaemi, S., Khanmohammadi, S., and Daneshvar, S. 2022. A hybrid high-order type-2 FCM improved random forest classification method for breast cancer risk assessment. *Applied Mathematics and Computation*, 424, 127038.



- Huljanah, M., Rustam, Z., Utama, S., and Siswantining, T. 2019, June. Feature selection using random forest classifier for predicting prostate cancer. In *IOP Conference Series: Materials Science and Engineering*, 546(5), 052031.
- Imran, B., Hambali, H., Subki, A., Zaeniah, Z., Yani, A., & Alfian, M. (2022). Data Mining Using Random Forest, Naïve Bayes, and Adaboost Models for Prediction and Classification of Benign and Malignant Breast Cancer. *Jurnal Pilar Nusa Mandiri*, 18(1), 37-46. <https://doi.org/10.33480/pilar.v18i1.2912>
- Islam, M., Haque, M., Iqbal, H., Hasan, M., Hasan, M., and Kabir, M.N. 2020. Breast cancer prediction: a comparative study using machine learning techniques. *SN Computer Science*, 1(5), 1-14.
- Jackins, V., Vimal, S., Kaliappan, M., and Lee, M.Y. 2021. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), pp 5198-5219.
- Jadhav, M., Thakkar, Z., and Chawan, P.M. 2019. Breast cancer prediction using supervised machine learning algorithms. *International Research Journal of Engineering and Technology (IRJET)*, 6.
- Jayaraj, D., and Sathiamoorthy, S. 2019, November. Random Forest based Classification Model for Lung Cancer Prediction on Computer Tomography Images. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 100-104.
- Kaur, P., Kumar, R., and Kumar, M. 2019. A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications*, 78(14), 19905-19916.
- Keleş, M.K. 2019. Breast cancer prediction and detection using data mining classification algorithms: a comparative study. *Tehnički vjesnik*, 26(1), 149-155.
- Khourdifi, Y., and Bahaj, M. 2018, December. Applying best machine learning algorithms for breast cancer prediction and classification. In *2018 International conference on electronics, control, optimization and computer science (ICECOCS)*, 1-5.
- Khoshgoftaar, T.M., Golawala, M., and Van Hulse, J. 2007, October. An empirical study of learning from imbalanced data using random forest. In *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, 2, 310-317.
- Kumar, A., and Poonkodi, M. 2019. Comparative study of different machine learning models for breast cancer diagnosis. In *Innovations in soft computing and information technology*, 17-25.
- Li, Y., and Chen, Z. 2018. Performance evaluation of machine learning methods for breast cancer prediction. *Appl Comput Math*, 7(4), 212-216.
- Macaulay, B.O., Aribisala, B.S., Akande, S.A., Akinnuwesi, B.A., and Olanjo, O.A. 2021. Breast cancer risk prediction in African women using Random Forest Classifier. *Cancer Treatment and Research Communications*, 28, 100396.
- Mashudi, N.A., Rossli, S.A., Ahmad, N., and Noor, N.M. 2021, March. Comparison on Some Machine Learning Techniques in Breast Cancer Classification. In *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 499-504.
- Mathew, T.E., 2019. 'A comparative study of the performance of different Support Vector machine Kernels in Breast Cancer Diagnosis'. *International Journal of Information and Computing Science (IJICS)*, 6(4), 432-441.
- Mathew, T.E., 2019, 'A Logistic Regression with Recursive Feature Elimination, for Breast Cancer Diagnosis', *International Journal on Emerging Technologies (IJET)*, 10(3), 55-63.
- Mathew, T.E., 2019, 'Simple and Ensemble Decision tree Classifier based detection of Breast Cancer', *International Journal of Scientific & Technology Research (IJSTR)*, 8(11), 1628-1637.
- Mathew, T.E., Anil Kumar, K.S., 2020, 'A Logistic Regression based hybrid model for Breast Cancer Classification', *Indian Journal of Computer Science and Engineering (IJCSE)*, 11(6), 899-903.
- Mathew, T.E., Anil Kumar, K.S., 2021, 'A Modified- Weighted- K -Nearest Neighbour and Cuckoo Search Hybrid Model for Breast Cancer Classification', *Indian Journal of Computer Science and Engineering (IJCSE)*, 12(1), 166-177.
- Napierała, K., 2012. Improving rule classifiers for imbalanced data. *Poznan University of Technology*.
- Paul, A., Mukherjee, D.P., Das, P., Gangopadhyay, A., Chintha, A.R., and Kundu, S. 2018. Improved random forest for classification. *IEEE Transactions on Image Processing*, 27(8), 4012-4024.
- Quist, J., Taylor, L., Staaf, J., and Grigoriadis, A. 2021. Random forest modelling of high-dimensional mixed-type data for breast cancer classification. *Cancers*, 13(5), p 991.
- Raj, S., Singh, S., Kumar, A., Sarkar, S., and Pradhan, C. 2021. Feature selection and random forest classification for breast cancer disease. *Data Analytics in Bioinformatics: A Machine Learning Perspective*, pp 191-210.
- Rohan, T.I., Siddik, A.B., Islam, M., and Yusuf, M.S.U. 2019. A precise breast cancer detection approach using ensemble of random forest with AdaBoost. In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 1-4.
- Shaik, A.B., and Srinivasan, S. 2019. A brief survey on random forest ensembles in classification model. In *International Conference on Innovative Computing and Communications*, 253-260.
- Shahhosseini, M., and Hu, G. 2020, August. Improved weighted random forest for classification problems. In *International Online Conference on Intelligent Decision Science*, 42-56.
- Sharma, S., Aggarwal, A., and Choudhury, T. 2018, December. Breast cancer detection using machine learning algorithms. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 114-118.
- Wang, S., Wang, Y., Wang, D., Yin, Y., Wang, Y., and Jin, Y. 2020. An improved random forest-based rule extraction method for breast cancer diagnosis. *Applied Soft Computing*, 86, 105941.
- Yap, B.W., Rani, K.A., Rahman, H.A.A., Fong, S., Khairudin, Z. and Abdullah, N.N., 2014. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, 13-22.
- Yifan, D., Jialin, L., and Boxi, F. 2021, May. Forecast Model of Breast Cancer Diagnosis Based on RF-AdaBoost. In *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 716-719.



- Zhu, M., Xia, J., Jin, X., Yan, M., Cai, G., Yan, J., and Ning, G. 2018. Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access*, 6, 4641-4652.
- Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., & Fan, J. (2020). Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*, 8, 96946-96954.
<https://doi.org/10.1109/ACCESS.2020.2993536>



RNI No. KERHIN/2017/70008

यू जी सी से अनुमोदित पत्रिका

ISSN No. 2456-625X



शोध सरोवर पत्रिका

10 अप्रैल 2022, वर्ष 6, अंक 22

'आरती', फोरेस्ट ऑफिस लेन, वषुतक्काट्टु, तिरुवनन्तपुरम -14

www.shodhsarovarpathrika.co.in

Peer reviewed Journal (पिअर रिव्यूड पत्रिका)



बहुचर्चित हिन्दी कथाकार श्रीमती मन्नू भंडारी (1931-2021)

त्रैमासिक हिन्दी शोध पत्रिका

अखिल भारतीय हिन्दी अकादमी, तिरुवनन्तपुरम, केरल राज्य।



शोध सरोवर पत्रिका

आरती, वषुतक्काट्ट, तिरुवनन्तपुरम - 695 014, केरल राज्य।

RNI No. KERHIN/2017/70008 ISSN No. 2456-625 X

वर्ष 6

अंक 22

त्रैमासिक हिन्दी शोध पत्रिका

10 अप्रैल 2022

इस अंक में

पीयर रिव्यू समिति:

डॉ.टी.के.नारायण पिल्लै

डॉ.शांति नायर

डॉ.के.श्रीलता

मुख्य संपादक

डॉ.पी.लता

प्रबंध संपादक

डॉ.एस.तंकमणि अम्मा

सह संपादक

प्रो.सती.के

डॉ.एस.लीलाकुमारी अम्मा

श्रीमती वनजा.पी

संपादक मंडल

डॉ.बिन्दु.सी.आर

डॉ.षीना.यू.एस

डॉ.सुमा.आई

डॉ.एलिसबत्त जोर्ज

डॉ.लक्ष्मी.एस.एस

डॉ.धन्या.एल

डॉ.कमलानाथ.एन.एम

डॉ.अश्वती.जी.आर

संपादकीय	:		3
मन्नू भंडारी का उपन्यास साहित्य	:	डॉ.पी.के.प्रतिभा	4
आधुनिक संतों की वैभव गाथा	:	गंगा कोइरी	9
‘संत न बाँधे गाँठड़ी’	:		
‘महाभोज’ में राजनीति	:	डॉ.सुमा.आई	15
यौन उत्पीड़न की सच्चाई का पर्दा उठाती	:	डॉ.बीना आर्या	17
सुभद्रा कुमारी चौहान की कहानियाँ	:		
मन्नू भंडारी की कहानी ‘सज़ा’ में न्याय	:	डॉ.एलिसबत्त जोर्ज	21
व्यवस्था की विसंगति	:		
‘सबसे बुरी लड़की’ की गवाही	:	डॉ.प्रकाश.ए	22
पारिवारिक विघटन की समस्याएँ	:	डॉ.षीबा शरत.एस	26
‘आपका बंटी’ में	:		
कोरोना काल में जड़ता के विरुद्ध सामर्थ्य	:	नेहा साव	28
और आत्मविश्वास की कविता ‘प्रार्थना’	:		
‘आपका बंटी’ उपन्यास में चित्रित	:	डॉ.धन्या.एल	32
बाल मनोविज्ञान	:		
‘आपका बंटी’ में चित्रित स्त्री-विमर्श	:	डॉ.रीनाकुमारी.वी.एल	33
इंदिरा गोस्वामी के उपन्यासों में मानवतावाद	:	मिनहाज अली	36
हिंदी कविता में कश्मीर की बहुरंगी	:	उमर बशीर	42
संस्कृति एवं समाज के विविध आयाम	:		
‘बिना दीवारों के घर’ नाटक में	:	डॉ.लक्ष्मी.एस.एस	46
स्त्री जीवन	:		

यू जी सी से अनुमोदित पत्रिका

शोध सरोवर पत्रिका, तिरुवनन्तपुरम, वर्ष 6 अंक 22 10 अप्रैल 2022



‘सबसे बुरी लड़की’ की गवाही

◆ डॉ. प्रकाश.ए

सार-डॉ.नीलमजी ने एक भुक्तभोगी गवाह की हालत के तौर पर अपनी सामाजिकराजनैतिक दृष्टि की अभिव्यक्ति दी है। ‘सबसे बुरी लड़की’ नामक संकलन की आत्मकथात्मकशैली की कविताएँ नागरिकों को अपने मूलाधिकार के प्रति सचेत करती हैं। कविताओं से संप्रेषित स्वानुभूति की प्रखरता प्रताड़ित नारियों को शोषणहीन जीवनपरिस्थिति के निर्माण की प्रेरणा देती है।

बीज शब्द - 1.रचनार्थमिता 2.प्रेरक तत्व 3.भुक्तभोगी 4.गवाही 5.विचार प्रणाली 6.संविधान 7.मूलाधिकार

जिनकी कलम को हमेशा दकियाकिनार करने का प्रयास किया गया है उनकेलिए सादर समर्पित कविता संग्रह का शीर्षक है ‘सबसे बुरी लड़की’, जिसमें युवा कवयित्री डॉ.नीलम की छोटी-सी छप्पन कविताएँ संग्रहीत हैं। हिंदी के दलित साहित्य की नई पीढ़ी के हस्ताक्षरों में डॉ.नीलम का स्थान विशेष उल्लेखनीय है कि अपनी रचनाओं के ज़रिए तत्कालीन समय और स्थिति के थप्पड़ से प्रताड़ित विद्रोही नारी के प्रतिनिधि बनना वे अपना हक समझती हैं।

रचनार्थमिता का आधार- जन्म लेकर जिस वातावरण में पलती और बढ़ती है वहाँ की विसंगतियों एवं विडंबनाओं को कविता की शकल के अंतर्गत आत्मकथन की शैली में प्रस्तुत करने की कोशिश सराहनीय है। हिंदी साहित्य से संबंधित उनकी कतिपय रचनाएँ प्रकाशित हो चुकी हैं। ‘सबसे बुरी लड़की’ संकलन की 11वीं कविता का शीर्षक है, जिसकी पंक्तियों के अध्ययन से ही डॉ.नीलम जी की रचनार्थमिता के प्रेरकतत्व की समझ पकड़ में आती है। थोपे गए सांस्कृतिक बिंबों पर प्रश्नचिह्न लगाने की प्रवृत्ति उनकी रचना की खासियत है। तथाकथित सभ्यता एवं कुलमर्यादा पर व्यंग्य कसती हुई उनका कहना है कि

मैं बहुत बुरी लड़की हूँ/क्योंकि मैं हँसती हूँ/
सभ्य लड़कियाँ हँसती नहीं/उनकी आँखों में
होती हैं/हया और शर्म/मेरी आँखों की/शर्म और
हया मर गयी है/क्योंकि मैं खुलकर हँसती हूँ
(सबसे बुरी लड़की-पृष्ठ 23)

‘रचना की राजनीति’ नामक लेख में मैनेजर पाण्डेय ने स्त्री-पराधीनता के विरुद्ध खड़ी होनेवाली मीराबाई का जिक्र करके नाभादास की इस पंक्ति को उद्धृत किया कि ‘लोकलाज कुलशृंखला तजि मीरा गिरिधर भजी’। आगे उन्होंने लिखा कि “अपने भारतीय समाज में अगर कोई स्त्री हँस दे, तो चार लोग एक साथ बोलेंगे कि अरे ये हँसती है। मतलब जिस समाज में स्त्रियों के होंठ खोलने का भी दायरा पहले से तय हो, उसमें स्त्री की पराधीनता कितनी जबर्दस्त है, यह हम सब जानते हैं (आलोचना में सहमति असहमति मैनेजर पाण्डेय, पृ.28)

पराधीनता एवं पारम्परिक बिम्बों के विरुद्ध- हिंदी में स्त्री जीवन की वास्तविकताओं का कलात्मक चित्रण करनेवाली लेखिकाओं एवं नारीमुक्ति के आंदोलन चलानेवाली कार्यकर्ताओं की कोई कमी नहीं है। सामंती संस्कृति के अवशेषों की कुलकानि की आड़ को छेदकर उन्मुक्त रूप से विद्रोहात्मक स्वर उठाने की कोशिश बहुत विरलों ने की है। डॉ.नीलमजी कविता की शकल में अपने जीवन परिवेश की संकीर्णता को आत्मविश्लेषण के तौर पर प्रस्तुत कर रही हैं, इसकेलिए उन्होंने किसी भी परम्परागत रचना-पद्धति को नहीं अपनाया। इसलिए कलावाद के हठधर्मी लोग, जो आर्तनाद में भी बाँसुरी का स्वर खोजते हैं, उनको तसल्ली देकर वे कहती हैं कि

मैं कोई कवि नहीं
जो अपने प्रतीकों और बिंबों से
सब कुछ बयान कर दूँ,
मैं कोई गीतकार नहीं

पाण्डेय, वाणी प्रकाशन, नई दिल्ली, 110002, प्रथम संस्करण, वर्ष-2013

(4) समकालीन हिन्दी कविता भाग 1- संदेवेंद्र भारती प्रखर, रश्मि प्रकाशन, लखनऊ पहला संस्करण, वर्ष-2020

(5) हिन्दी आलोचना की पारिभाषिक शब्दावली- डॉ.अमरनाथ, राजकमल प्रकाशन, नई दिल्ली, पाँचवाँ छत्र संस्करण, वर्ष-2018.

(6) आधुनिक हिन्दी आलोचना के बीज शब्द- बच्चन सिंह, वाणी प्रकाशन, नई दिल्ली, 110002, आवृत्ति संस्करण, वर्ष 2015

◆ असिस्टेंट प्रोफेसर
हिन्दी विभाग, सरकारी कॉलेज
कार्यवट्टम, केरल राज्य।
मो-9446413027

(पृ.सं.21 के आगे)

समाने लगा कि अब शायद पापा कभी प्यार नहीं करेंगे। और सचमुच; उसके बाद मैंने कभी उनका प्यार नहीं पाया आज तक नहीं।”¹ मन्नु जी ने पूरी कहानी में पन्द्रह साल की उस बच्ची की आशा-निराशाओं, दुख-दर्दों, कुंठा-तनावों की बारीकी से चित्रण किया है।

वर्तमान युग की यही विशेषता है कि अपराधी को छूट मिल जाती है, समाज में उसकी प्रतिष्ठा होती है; लेकिन निरपराधी को सज़ा ही प्राप्त होता है, उसके मान-सम्मान का कोई मूल्य नहीं होता। कान्ता मामा का कथन एकदम सही है कि “आज के ज़माने में तो गुनहगार अपने को साफ बचाकर ले जाते हैं। लाखों हज़म करके मूँछों पर ताव देते घूमते हैं। फाइल की फाइलों को गायब करवा देते हैं। और एक ये है कि बिना गड़बड़ किए जेल भोग रहे हैं।”² जब व्यक्ति अपराध किये बिना ही अपराधी बन जाता है तो उसकी सद्भावना भी बदलने लगती है। यह एक मनोवैज्ञानिक तथ्य है। ऐसे ही समाज में गुनहगार बनता है। भारी यंत्रणाओं से गुज़र जाने पर आशा सोचने लगती है- “इससे तो पापा सचमुच ही ऑफिस का रुपया मार लेते तो अच्छा होता। - ईमानदारी करके ही कौन बड़ा सुख मिल रहा है।”³

कानूनी प्रक्रिया में होनेवाला विलंब बेकसूर के लिए किसी सज़ा से कम नहीं है। न्याय में देरी अन्याय ही होता है। इस कहानी में प्रतिपादित हालत केवल एक व्यक्ति या परिवार तक सीमित नहीं होती। यह कोई काल्पनिक घटना भी नहीं है। हमारे देश में चारों ओर ऐसी घटनाएँ होती रहती हैं। न्यायालयों में विलंबित मुकदमों की संख्या दिन पर दिन

बढ़ती जा रही है। भारत में न्यायाधीशों की कमी और मुकदमों की बढ़ती संख्या के कारण या फिर न्यायालयों की ढीली कार्य-प्रणाली की वजह से ऐसी स्थिति पैदा होती है। इसके अलावा तकनीकी का अभाव, अधिवक्ताओं द्वारा मामले को अटके रखने के लिए जानबूझकर विलंब कराना, सुनवाई के बीच लंबे अवकाश लेने की प्रथा, न्याय विभाग में संचार की कमी आदि भी इस विलंब के कारण बनते हैं। सन् 2012 में घटित निर्भया केस के विचारण में 7 वर्ष लग गये थे। ऐसे कई उदाहरण ले सकते हैं। इस समस्या का हल निकालना आवश्यक है। इसके लिए, मुकदमे की सुनवाई में कुछ समय सीमा निर्धारित करना अनिवार्य होगा। अमेरिका में यह समय सीमा 3 वर्ष निर्धारित हुआ है। मन्नु भंडारी ने प्रस्तुत कहानी ‘सज़ा’ के द्वारा राष्ट्र की एक ज्वलंत समस्या की ओर पाठकों का ध्यान आकृष्ट किया है। उन्होंने यह साबित किया है कि महिला कहानीकार भी समाज की समस्याओं के प्रति सतर्क है और उसे सफलतापूर्वक प्रस्तुत कर सकती हैं।

संदर्भ

1. मन्नु भंडारी, यही सच है (कहानी संकलन)- सज़ा- पृ.61
2. वही, पृ.60
3. वही, पृ.63

◆ असिस्टेंट प्रोफेसर
सरकारी वनिता कॉलेज
तिरुवनंतपुरम, केरल राज्य।
मो-9495606105

शोध सरोवर पत्रिका

Shodh Sarovar Pathrika

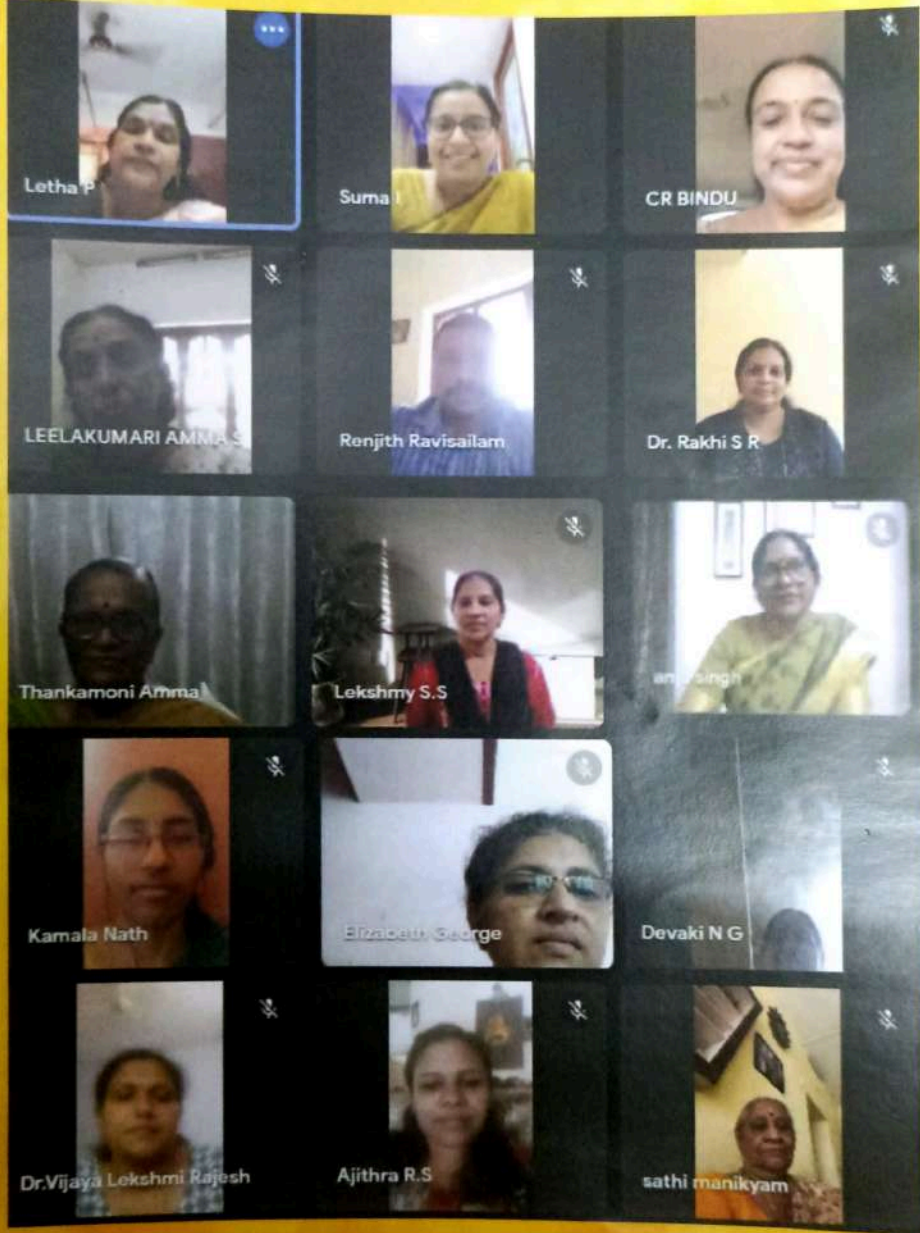
(त्रैमासिक हिन्दी शोध पत्रिका / Quarterly Hindi Research Journal)

RNI No. KERHIN/2017/70008 ISSN No. 2456-625X

Rs. 100/-

10 अप्रैल 2022

Volume 6, Issue 22, Thiruvananthapuram



'अखिल भारतीय हिन्दी अकादमी' के अंतर्राष्ट्रीय महिला दिवस समारोह का श्रीमती अंजु सिंह (सहायक निदेशक, केन्द्रीय हिन्दी निदेशालय) द्वारा उद्घाटन।

Printed & Published by Dr.P.Letha (on behalf of Akhil Bharatheeya Hindi Academy),
Arathi, T.C. 14/1592, Forest Office Lane, Vazhuthacaud, Thiruvananthapuram -14,
Printed at Abi Design & Pre-Press, Karumom, Thiruvananthapuram -2 & Edited by Dr. P. Letha