



## EFFECT OF ZYCOTHERM ON PROPERTIES OF POROUS FRICTION COURSE MIXES

Arijit Kumar Banerji<sup>1\*</sup>, Md. Hamjala Alam<sup>2</sup>, Antu Das<sup>3</sup>

<sup>1</sup>Department of Civil Engineering, Dr. B. C. Roy Engineering College, Durgapur, WestBengal, India. Email: [arijit.banerji@bcrec.ac.in](mailto:arijit.banerji@bcrec.ac.in),

<sup>2</sup>Department of Civil Engineering, Dr. B. C. Roy Engineering College, Durgapur, WestBengal, India. Email: [mdhamjala.alam@bcrec.ac.in](mailto:mdhamjala.alam@bcrec.ac.in)

<sup>3</sup>Quality Assurance and Control Engineer, B.S. GeoTech Pvt. Ltd., Konnagar, WestBengal, India., Email: [antudas94@gmail.com](mailto:antudas94@gmail.com)

**Abstract** — Warm mix asphalt (WMA) is a promising green technology with the potential to replace hot mix asphalt (HMA) in road construction. The purpose of this research is to assess the efficacy of ZycoTherm WMA green technology in the construction of porous pavement. Porous pavement mixes comprised of open graded porous friction course materials with large voids reduce pressure on urban drainage systems and enhance natural water supply during rainstorms. Using WMA additive ZycoTherm and VG 30 grade bitumen, this research focused on the properties of PFC grading bituminous mixes specified by the Indian Road Congress code IRC: 129-2019. ZycoTherm is an eco-friendly nano-organic anti stripping additive that allows for lower mixing temperatures. Volumetric properties, water drain-down, permeability, and un-aged and aged abrasion resistance tests were performed on porous paving mixes with and without ZycoTherm. This study examines PFC mixtures with bitumen contents of 5.0 to 6.0 percent. The experimental results were analysed in order to determine the most vital controlling parameters. According to the results, ZycoTherm shows the potential to improve the performance properties of porous mixes.

**Keywords** — *warm mix asphalt, hot mix asphalt, ZycoTherm, porous pavement.*

### 1. INTRODUCTION

Porous friction course mixes are a type of hot mix asphalt that is distinguished by a large percentage of inter-connected air voids, coarse granular grading, and suitable stone-on-stone contact. Bituminous PFC are also known as Open Graded Friction Courses (OGFC), Porous Asphalt (PA), Permeable Friction Courses (PFC), Open Graded Porous Asphalt (OGPA), Open Graded Asphalt Mixes (OGA) by various agencies around the world [1-3]. Between 19 and 50 mm thick, PFC layers have a minimum air void content of 18 % [4]. High air void content makes the PFC mixes permeable, which reduces the pressure on urban drainage systems and facilitates natural water supply in case of a rainstorm. Because of its porous structure, PFC can also be used as a storm water filter in addition to its other functions. PFC implements a uniform grading of aggregates in order to achieve the high proportion of air voids required. According to the research, PFCs provide a variety of benefits, including improved skid resistance, reduced pavement noise levels, and enhanced night time visibility during wet weather conditions, in addition to hydroplaning prevention [5-7]. The fact that many of these characteristics are in conflict with one another is obvious.

\*Corresponding author: Arijit Kumar Banerji, [arijit.banerji@bcrec.ac.in](mailto:arijit.banerji@bcrec.ac.in)

Unfortunately, the load bearing capacity of such PFC pavements is less than conventional pavements due to the absence of the fine particles [8]. It also weakens the adhesion bond between the binder and aggregate surface, resulting in an increased possibility of cracking, ravelling, and pothole formation. The moisture damage properties of PFC mixtures are crucial since they are expected to perform hydrological activities and are in continual exposure to water. In recent years, a significant number of porous asphalts have been developed with the goal of balancing these various requirements. In addition to providing a number of other advantages, the aggregate gradation of a bituminous PFC mixture has an impact on the ability of the mixture to withstand traffic loading. The effects of aggregate gradation on the properties of PFC were investigated for this study in order to evaluate volumetric and performance characteristics.

WMA is a recognised pavement construction technology that addresses environmental and economic concerns. In the road construction sector, WMA has the potential to substitute the practise of HMA [9]. The goal of WMA is to manufacture paving mixes at temperatures lower (20–40°C) than those used in hot mix. WMA is a technology that aims to minimise the amount of energy used and the amount of emissions produced during asphalt paving. The application of WMA technologies in asphalt paving can reduce energy usage by 40%, resulting in a reduction in greenhouse gas emissions [10]. Workers in the manufacturing plant and on the worksite are inhaling much fewer plumes of smoke. This also enables extended hauling distances, minimises the possibility of compaction difficulties, and necessitates less time for the compacted mixtures to cool before they are opened to traffic [11]. Currently, WMA technologies are classified into three broad categories: (i) foaming techniques, (ii) organic additive use, and (iii) chemical additive. In recent years, several articles on WMA technologies have discussed the usage of various additives. To test the resistance of WMA porous mix with Sasobit wax additive, Hamzah et al. used indirect tensile strength ratio test [12]. The results showed that dynamic water movement considerably reduced stripping resistance. Wurst and Putman compared Evotherm and foamed WMA with HMA in a warm-mix environment [13]. They discovered that by employing the WMA technologies, fibres from OGFC mixes can be removed when comparing three major criteria: drain down, permeability, and abrasion resistance. According to Bairgi et al., compaction and long-term ageing both contribute to the promotion of rutting and stripping parameters [14]. WMA and OGFC mixtures comprising RAP and PMB was analyzed by Ingrassia et al. [15] using an in-situ falling weight deflectometer. They found that the OGFC and WMA combinations outperformed the HMA mixture in terms of reliability and durability. However, the differences in aggregate gradations, bitumen sources, mix designs, and testing conditions in India, in comparison to other countries, require specific studies to gain an understanding of the behaviour of WMA mixes. Warm mix PFC has been evaluated in limited performance trials with WMA technologies in recent years. Furthermore, because of the above WMA properties, ravelling and drain down could be reduced by using warm mix PFC, especially when anti-stripping WMA additives are utilized.

## **2. OBJECTIVES AND SCOPE**

The purpose of this study is to assess the practicality of the application of the ZycTherm warm mix additive to PFC mixes. PFCs have not yet been tried by Indian authorities in national highways or expressways, but recently 2019 specifications have been recommended. PFC mixes with gradations suggested by the Indian Roads Congress (IRC: 129-2019) [16] specifications for open-graded friction courses were explored in this study. The performance and reliability characteristics of PFC combinations were studied by taking into account the following mix design qualities: volumetric properties, porosity, drain down, un-aged and aged abrasion loss. The Marshall approach is used to compact the mix, which will consist of 25

blows on each side of the mixture. The experimental data was analysed to determine the most significant influencing parameters and their relative importance.

### 3. MATERIAL DESCRIPTION AND MIX PREPARATION

#### A. Materials Used

The investigation focused on the PFC mixes that corresponded to the gradation specified by the IRC: 129-2019 requirements, as shown in Table 1. A grading specification that was closer to the middle of the range was chosen [17]. Portland Pozzolana cement was employed as the filler fines. In accordance with Indian Standards and ASTM test procedures, the physical properties of aggregate and bitumen were determined. Detailed results of the tests are reported in Table 2. Additionally, each mixture includes WMA ZycoTherm at 0.1% by weight of bitumen. It is an innovative, high performance anti stripping additive for asphalt mixes that, in comparison with conventional anti stripping additives, chemically bonds with the aggregates to generate a flawless barrier that lasts for an extended period of time. It enables 100% coating and wetting of the pavement, as well as improved anti-stripping capabilities, resulting in increased moisture resistance of the pavement [18,19]. When compared to conventional additives, ZycoTherm modified bitumen coats aggregate surfaces substantially faster and with less effort than other types of bitumen. Its addition has the effect of lowering the temperature at which asphalt mixtures are produced and compacted. ZycoTherm is the chemical WMA additive with the following properties given in the Table 3 [18]. The bitumen was first heated to 140°C for mixing.

Table 1 - IRC: 129-2019 Aggregate Grading

IS Sieve Size (mm)	13.2	9.5	4.75	2.36	0.075
% Passing	100	92.5	30	7.5	3

Table 2 – Physical Properties of Coarse Aggregate and Bitumen

Materials	Properties	Results	Standards
Aggregates	Aggregate Impact Value, %	12.28	IS 2386 (P4)
	Combined Flakiness and Elongation Index, %	24.67	IS 2386 (P1)
	Fractured Particles, %	97	ASTM D8521
	Water absorption, %	0.55	ASTM C127
	Specific gravity (bulk)	2.82	IS 2386 (P3)
	Specific gravity (apparent)	2.85	
	Los Angles Abrasion, %	19.34	ASTM C131
Bitumen	Stripping Test	100	
	Specific gravity at 27°C	1.02	IS 1202
	Penetration at 25°C, 100 g, 5 s	57	IS 1203
	Softening point, (R&B), °C	48.5	IS 1205
	Ductility at 27°C, cm	>100	IS 1208
	Flash Point, °C	310	IS 1209

Table 3 - Physical Properties of WMA Additive ZycoTherm [18]

Property	Descriptions
Specific gravity	0.97
Viscosity	100-500 cps
Flash point	>80°C
Colour	Pale yellow
pH value	10% soluble in neutral water
Physical state	Liquid

## B. Bituminous PFC Mix Preparations

Asphalt mixes are made by first preheating aggregates at mixing temperature in oven before mixing them together. The compaction temperature was maintained in the oven for one hour after mixing. For HMA mixes, 165°C mixing temperatures was used, and for WMA mixes, 145°C mixing temperatures was adopted. Compaction was performed at a temperature that was 10 degrees Celsius lower than mixing. All of the PFC mixtures were made into standard specimens of 100 mm diameter by 63.5 mm height. 5.0 percent, 5.50 percent, and 6.0 percent bitumen were used in the design trials. 25 blows of the Marshall Hammer was applied to each side, and the samples were let cool for 24 hours at room temperature before being investigated in accordance with IRC 129: 2019 [16].

## 4. PFC PERFORMANCE TESTS

### A. Volumetric Properties of Compacted Specimens

The service life of PFC pavements varies depending on the environment, vehicular traffic and loads, as well as the construction and design standards used to construct them. The minimal bitumen content at which the PFC mixes are capable of meeting ASTM D7064 [20] design standards was chosen as the optimum bitumen content (OBC). The volumetric properties were calculated using the bulk specific gravity and maximum specific gravity results. The theoretical maximum density of the loose mix was determined using effective specific gravity as per standards of ASTM D 2041 [21]. When the maximum specific gravity and bulk specific gravity were obtained, the air void was calculated by multiplying the two quantities. Percentage of voids in compacted mix ( $VCA_{mix}$ ) and percentage of voids in dry-rodded state ( $VCA_{drc}$ ) were used to detect whether there was stone-on-stone (SoS) contact in the PFC mix gradations. The density of a specific coarse aggregate sample is determined by conducting a dry rodded test. Using a constant volume cylinder filled with three layers and 25 tamps per layer, this test provides an understanding of the internal arrangement of aggregates in a dry state. PFC mix SoS contact condition was approved by examining the  $VCA_{mix}$  to  $VCA_{drc}$  ratio. All of the mixtures that corresponded to the gradation were observed to have mean void percentage values greater than 18 percent. Table 4 shows the results of these studies for maximum specific gravity, bulk specific gravity of PFC mix,  $VCA_{mix}/VCA_{drc}$ , and void percentage for the purposes of comparison.

Table 4 - Volumetric Properties of PFC mix

Property	Results		
BC (%)	5.0	5.5	6.0
Air Void, %	21.79	21.44	19.58
VCA <sub>mix</sub>	37.4	37.6	36.5
VCA <sub>drc</sub>	44.1		
VCA <sub>mix</sub> / VCA <sub>drc</sub>	0.84	0.85	0.82

## B. Permeability

Permeability or hydraulic conductivity, refers to the rate at which water passes through a material. Drainage properties are critical to the performance of a PFC mix. Pore size measurements help determine the drainage characteristics of these mixes. The contact surface of the Marshall sample with mould was double coated to avoid water leakage. The time it took for the water to descend from initial head (h1) to final head (h2) was recorded in seconds. Equation 1 was used to compute the permeability coefficient (k) and the results ranged from 92 to 140 m/day. A permeability rate of 100 m/day was advised by Mallick and Chen to provide adequate performance [22,23]. Because permeability coefficient value should be greater than 100 metres per day for proper drainage, this mix's bitumen concentrations of 5.0 and 5.5 percent met this requirement. Permeability was shown to be closely linked to air-void trends, which is generally believed to be true.

$$K(\text{m/day}) = \frac{(3600 \times 24) al}{1000 At} \ln \frac{h_1}{h_2} \quad (1)$$

where; l is sample thickness in mm, a is standpipe cross-sectional area in mm<sup>2</sup>, A is sample cross-sectional area in mm<sup>2</sup> and K is permeability constant. Table 5 summarises the permeability data for each of the PFC combinations.

Table 5 - Permeability and Air Void results

Property	Results		
BC (%)	5.0	5.5	6.0
K(m/day)	140	133	92
Air Void (%)	21.79	21.44	19.58
K(m/day)	Adding 0.1% ZycTherm@5.5%: 178		
Air Void (%)	21.82		

## C. Cantabro Abrasion Loss

The compacted specimens were subjected to the Cantabro abrasion test in terms of determining their durability. In the un-aged state, the abrasion loss should not exceed 20%, and in the aged state, it should not exceed 30%. For the mix design of PFCs, most agencies now suggest this test as either a mandatory or optional test. The abrasion loss on un-aged and aged samples are denoted as UAL and AAL, respectively.



Figure 1 - (a) Specimen in abrasion cylinder (b) Fully compacted specimen and after abrasion.

The un-aged abrasion loss test was performed on a Los Angeles abrasion drum (Figure 1) at 30 to 35 revolutions per minute for 300 revolutions. Samples of asphalt mixtures were tested by determining how much weight were reduced when compared with their initial weight. As a result, the bitumen in PFC hardens faster resulting in a loss of cohesive and adhesive strength, which can lead to ravelling. Consequently, an accelerated ageing test was carried out on the mix design to see how it performs at 85°C for 120 hours. The test is performed according ASTM D7064 [20] in both aged and un-aged conditions. Tests will be performed on the specimens after they have been cooled to 25°C and stored for four hours. Individual sample abrasion loss should not above 50 percent, whereas UAL and AAL sample abrasion loss should not exceed 20 percent as well as 30 percent, respectively. The oxidation of bitumen makes it more stiff, which makes the bond between aggregate particles and the bitumen more brittle, which makes abrasion loss more likely in aged samples. The variation of UAL and AAL of HMA and WMA mix are shown in Table 6.

Table 6 - Unaged Abrasion Loss and Aged abrasion Loss results

Property	Results			
BC (%)	5.0	5.5	WMA	6.0
Unaged Abrasion Loss (%)	16.81	13.78	22.51	13.45
Aged Abrasion Loss (%)	10.85	10.70	14.55	8.92

#### D. Drain Down Test

Draindown testing was performed on un-compacted specimens in accordance with AASHTO T305 [24] and ASTM D6390 [25] standards. The test allows for the evaluation of an asphalt mixture's drain down potential at the stages of mixture design and field manufacturing, respectively. A thin bitumen coating is left behind after draindown, which is insufficient to prevent particles from being dislodged by traffic. It is also possible for the thin bitumen film to age more quickly and become brittle, thus worsening the ravelling problem. Increased bitumen content in PFC mixtures will have a greater chance of causing bitumen to drain away during transportation. This will result in dispersion of bitumen and improper bituminous mix, which will cause ravelling to occur.

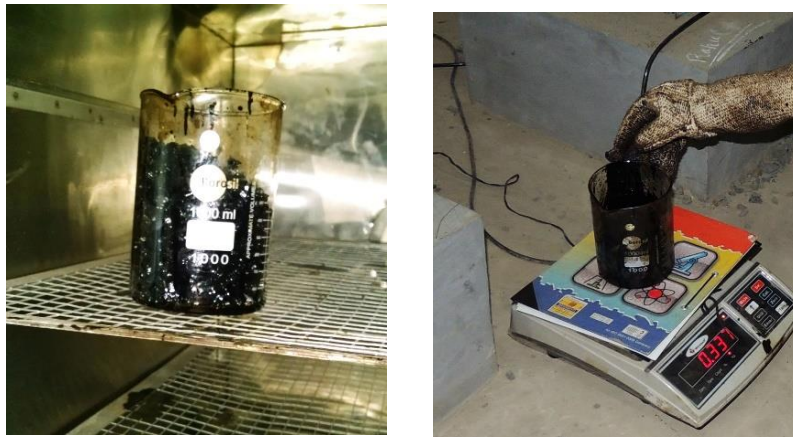


Figure 2 - (a) Specimen conditioning in oven (b) Flask with drained material.

The Schellenberg Binder Drainage Test was used to measure the amount of bitumen drained in this investigation. This experiment used glass jars with a diameter of 98mm and a height of 136mm (Figure 2). The sample is kept in glass jars for an hour at a temperature 15°C above mixing and compaction. After a 60 minutes, the jar is tipped to measure the remaining material. The Drain Down value should be within 0.3 percent of each other when calculated using Equation 2. Table 7 shows the drain down values for different bitumen contents.

$$\text{Drain down (\%)} = \frac{(\text{Final Weight of flask with Drain sample} - \text{Initial Weight of flask})}{\text{Initial Weight of flask}} \times 100 \quad (2)$$

Table 7 - Drain down Test results

Property	Results			
BC (%)	5.0	5.5	WMA	6.0
Drain down (%)	0.20	0.21	0.19	0.24

#### 4. CONCLUSION

The key objective of this study was to assess the practicability of using WMA technologies to produce acceptable PFC mixtures. This study examined ZycoTherm WMA mixes with HMA PFC based on draindown, permeability, and abrasion resistance as the key criteria. There were two asphalt mixtures examined and tested: one with 0.1 percent ZycoTherm, and the other with VG bitumen (HMA mix). For this experiment, the mixing and compacting temperatures were varied so that the performance of the mixes could be evaluated. First, volume and physical measurements were taken, followed by Cantabro, drain down, and permeability tests.

The stone-on-stone contact criterion was met by this gradation. They have good volumetric and permeability qualities for PFC mixtures compacted with 25 blows. Compared to PFC samples, WMA samples have higher water permeability ratings because they contain more air. Mixes with a bitumen concentration of 5.0 and 5.5 percent had air voids content of approximately 20 percent. Both the Unaged Abrasion Loss and Aged abrasion Loss values lies inside the 20% and 30% thresholds, respectively. Traditional PFC specimens are slightly more

susceptible to compaction than Zycotherm specimens, perhaps due to the lower manufacturing temperature. Mixes treated with WMA Zycotherm reduced drain down losses to less than 0.20%. According to these data, WMA technology may be used to lay porous asphalt at a lower temperature without a substantial deterioration. However, further research is required to better understand these mixtures.

**Acknowledgement:** The research is based upon the work supported by Dr. B. C. Roy Engineering College, Durgapur.

## References

- [1] Shankar, A. U., Suresha, S. N., & Saikumar, G. M. V. S., "Properties of porous friction course mixes for flexible pavements", *Indian Highways*, 42(3), 2014.
- [2] Suresha, S.N, Varghese George and Ravi Shankar, A.U., "Investigation of Porous Friction Courses (PFC) and Mixes: a Brief Overview", *Indian Highways*, Vol. 35, No.7, p. 21-43, 2007.
- [3] Banerji, A. K., Das, A., Mondal, A., Obaidullah, M., Biswas, R., & Saha, U., "Performance Assessment of Porous Friction Course (PFC) Mix modified with Cement as Filler Material", *American International Journal of Research in Science, Technology, Engineering & Mathematics*, 15-532, p. 83-88, 2015.
- [4] Mansour, T. N., & Putman, B. J., "Influence of aggregate gradation on the performance properties of porous asphalt mixtures", *Journal of Materials in Civil Engineering*, 25(2), 281-288, 2013.
- [5] Huber, G., "Performance Survey on Open-Graded Friction Course Mixes." *Synthesis of Highway Practice 284*, National Cooperative Highway Research Program, Transportation Research Board, 2000.
- [6] Kandhal, P. S. "Design, construction and maintenance of open-graded asphalt friction courses", *National asphalt pavement association information series 115*. America.
- [7] Alvarez, A. E., Martin, A. E., & Estakhri, C. "A review of mix design and evaluation research for permeable friction course mixtures", *Construction and Building Materials*, 25(3), 1159-1166, 2011.
- [8] Kandhal, P. S., & Mallick, R. B. "Open-Graded Asphalt Friction Course: State of the Practice. National Center for Asphalt Technology Report", (98-7), 1998.
- [9] Prowell, B. D., Hurley, G. C., & Crews, E., "Field performance of warm-mix asphalt at the NCAT test track", 86th Annual Meeting of the Transportation Research Board, Transportation Research Board, 2007.
- [10] Vaitkus, A., Čygas, D., Laurinavičius, A., & Perveneckas, Z., "Analysis and evaluation of possibilities for the use of warm mix asphalt in Lithuania", *Baltic J. Road Bridge Eng.*, 4(2), 80-86, 2009.
- [11] M. Zaumanis, *Warm Mix Asphalt Investigation Master of Science Thesis*, Technical University of Denmark, 2010.
- [12] M.O. Hamzah, M.Y. Aman, Z. Shahadan, M.H.M. Rosli, "Laboratory Evaluation of the Dynamic Stripping Test On Porous Asphalt Incorporating Sasobit", *7th International Conference on Road and Airfield Pavement Technology*, Bangkok, 2011, p. 130-141.
- [13] Wurst III, J. E., & Putman, B. J., "Laboratory evaluation of warm-mix open graded friction course mixtures", *Journal of Materials in Civil Engineering*, 25(3), 403-410, 2013.
- [14] Bairgi, B. K., Tarefder, R. A., & Ahmed, M. U., "Long-term rutting and stripping characteristics of foamed warm-mix asphalt (WMA) through laboratory and field investigation", *Construction and Building Materials*, 170, 790-800, 2018.
- [15] Ingrassia, L. P., Cardone, F., Ferrotti, G., & Canestrari, F., "Monitoring the evolution of the structural properties of warm recycled pavements with Falling Weight Deflectometer and laboratory tests", *Road Materials and Pavement Design*, S69-S82, 2021.
- [16] IRC: 129 Specifications for Open-Graded Friction Course. IRC, New Delhi, India, 2019.
- [17] Banerji, A. K., Das, A., Mondal, A., Biswas, R., & Obaidullah, M., "A Comparative evaluation on the properties of HMA with variations in aggregate gradation of laboratory and field produced mixes. *Int. J. Eng. Res. Technology*, p. 441-448, 2014.
- [18] Zydex, ZYCOTHERM Material Safety Data Sheet. Gotri, Vadodara:Zydex, Editor, 2014.
- [19] Tripathi, P., Ray, D. S., "Optimization of Different Anti Stripping Agents in Construction of Flexible Pavement: *International Journal of Engineering Research & Technology*, 9(7), p. 670-681, 2020.
- [20] ASTM D 7064, Standard practice for open graded friction course (OGFC) mix design, ASTM International, ASTM International, West Conshohocken, PA, 2013.
- [21] ASTM D 2041, Standard test method for theoretical maximum specific gravity and density of bituminous paving mixtures, ASTM International, West Conshohocken, PA, 2011.
- [22] Mallick, R.B., Kandhal, P.S., Cooley, L.A. Jr., and Watson, D.E., "Design, Construction, and Performance of New-Generation Open-Graded Friction Courses", *J. Association of Asphalt Paving Technologists*, 69, 391-423, 2000.



- [23] Chen, J.S., Lin, K.Y., and Young, S.Y., "Effects of Crack Width and Permeability on Moisture-Induced Damage of Pavements", J. Materials in Civil Engineering, 16(3), 276-282, 2004.
- [24] AASHTO T305. Standard Method of Test for Determination of Draindown Characteristics in Uncompacted Asphalt Mixtures, American Association of State Highway and Transportation Officials, Washington, D.C., 2001.
- [25] ASTM D 6390. Standard Test Method for Determination of Draindown Characteristics in Uncompacted Asphalt Mixtures, West Conshohocken, P.A., 2005.



**Arijit Kumar Banerji** is a B.Tech in Civil Engineering from BIET Suri and an M.Tech in Transportation Systems Engineering from IIT Guwahati. Since July, 2014 he is working as Assistant Professor at Dr. B.C Roy Engineering College, Durgapur, West Bengal. Presently he is also pursuing his Phd at NIT, Durgapur as part-time research scholar. He is having more than seven years of teaching experience. His areas of interests are Pavement material characterization, Pavement Design and Finite Element Modelling of pavement.



**Md. Hamjala Alam** is a B.Tech in Civil Engineering from Narula Institute of Technology, Kolkata and an M.Tech in Environmental Science & Engineering from Indian School of Mines Dhanbad. Since August, 2014 he is working as Assistant Professor in Civil Engineering at Dr. B.C Roy Engineering College, Durgapur, West Bengal. His areas of interests are Water Quality, Air and Noise Pollution and Control, Soil Stabilization, Soil Strength enhancement, Concrete properties and Pavement Design.



**Antu Das** is a B.Tech in Civil Engineering from Dr. B. C. Roy Engineering College, Durgapur and currently working as Quality Assurance and Control Engineer in B.S. GeoTech Pvt. Ltd., Konnagar, WestBengal. He has published two international journal paper and his areas of interests are highway design, soil investigation and behaviour of structures.



BEST  
BCREC Engineering & Science Transaction  
Journal website: [www.bcrecjournal.org](http://www.bcrecjournal.org)



# A NEW HYBRID EVOLUTIONARY ALGORITHM FOR OPTIMAL LOCATION OF SHUNT FACTS CONTROLLER TO OPTIMIZE TRANSMISSION LOSS, VOLTAGE DEVIATION AND VOLTAGE STABILITY

Susanta Dutta<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Dr. B C Roy Engineering College, Durgapur, West Bengal, India

**Abstract** - This paper presents a new hybrid chemical reaction optimization (HCRO) approach to find the optimal placement of Static Synchronous Compensator (STATCOM) to achieve optimal performance of power system. Optimal reactive power dispatch (ORPD) has a significant influence on optimal operation of power systems. However, getting optimal solution of ORPD is a cumbersome task for the researchers. The inclusion of flexible AC transmission system (FACTS) devices in the power system network for solving ORPD adds to its complexity. This paper presents HCRO for optimal allocation of STATCOM to minimize the transmission loss, improve voltage profile and voltage stability in a power system. It is found that the results obtained by the HCRO technique are superior to those obtained by other algorithms.

**Keywords:** *Chemical reaction optimization; Static Synchronous Compensator (STATCOM); Optimal Reactive Power Dispatch (ORPD)*

---

Corresponding Author

Contact address: Dr. B C Roy Engineering College, Fuljhore; Jemua Road; Durgapur-713206, West Bengal, India.  
email-address: [susanta\\_dutta@bcrec.ac.in](mailto:susanta_dutta@bcrec.ac.in)

<b>Nomenclature</b>	<i>rand</i> – random number
$Y_p'$ – Admittance of parallel component	$C_r$ – Crossover probability
$G_p', B_p'$ -Conductance and Susceptance of the parallel component respectively	$rand_{i,j}$ – random number between [0, 1]
$\theta_{jk}$ - Admittance angle of transmission line connected between jth and kth bus respectively	NL-number of load bus
$\delta_p'$ – Voltage source phase angle of the parallel components, respectively of STATCOM based facts devices	NG-number of generator bus
$E_p'$ -voltage source phase angle of the parallel converters, respectively of STATCOM based facts devices	NC-number of shunt compensator
$V_j, V_k$ – Voltage magnitude converters at the jth and kth bus respectively	NB-number of buses
$P_j, Q_j$ –Active and reactive power generation of the jth bus	$X_j^{\min}, X_j^{\max}$ -Maximum and minimum limits, respectively, of the jth variable
$Y_{jk}$ – Admittance of the transmission line connected between jth and kth bus	$X_{ij}$ – jth independent variable of the ith vector
$P_{gi}, P_{lj}$ – Active power generation and demand of jth bus	$E_p^{ms_1}$ – Potential energy of the new molecule.
$Q_{gi}, Q_{lj}$ – Reactive power generation and demand of jth bus	$E_k^{ms}, E_p^{ms}$ – Kinetic energy and potential energy, respectively, of the original molecule.
	<b>Abbreviation list</b>
	OPF- Optimal power flow
	GA- Genetic Algorithm
	PSO-particle swarm optimization
	IGA- immune GA
	IPSO- immune PSO
	HCRO- hybrid chemical reaction optimization
	CRO- chemical reaction optimization
	DE- differential evolution
	STATCOM- static synchronous compensator

$q_{cj}^{\min}, q_{cj}^{\max}$ – Reactive power injection limits of the jth shunt compensator	FACTS- flexible AC transmission systems HIA- hybrid immune algorithm
$t_j^{\min}, t_j^{\max}$ – Tap setting limits of the jth regulating transformer	LP- linear programming NLP- non-linear programming
$S_{lj}, S_{lj}^{\max}$ – Apparent power flow and maximum apparent power flow limit of the jth branch	MILP- mixed integer linear programming EP- evolutionary programming
$V_{lj}^{\min}, V_{lj}^{\max}$ – Minimum and maximum voltage limits, respectively, demand of the jth bus	SA- simulated annealing TS- tabu search
$P_{gj}^{\min}, P_{gj}^{\max}$ – Minimum and maximum active power generation limits, respectively, of the jth generator bus	ABC- artificial bee colony HSA- harmony search algorithm
$V_{gj}^{\min}, V_{gj}^{\max}$ – Minimum and Maximum voltage limits, respectively, of the jth generator bus	SVC-static VAR compensator ORPD- optimal reactive power dispatch
$Q_{gj}^{\min}, Q_{gj}^{\max}$ – Minimum and maximum reactive power generation limits, respectively, of the jth generator bus	NTVAC- non-linear time-varying acceleration coefficients PE- potential energy
$G_m$ – Conductance of the mth line connected between jth and kth buses	KE- kinetic energy
$h_{jk}$ – Admittance angle of transmission line of jth and kth bus respectively	

## 1. INTRODUCTION

In recent years, the electrical power demand has grown rapidly. The electric power grid is the largest man-made machine in the world. It consists of synchronous generators, transformers, transmission lines, switches and relays, active/reactive components, and loads. Power system networks are complex systems that are nonlinear, non-stationary, and prone to disturbances and faults. Reinforcement of a power system can be accomplished by improving the voltage profile, increasing the transmission capacity and others. Nevertheless, some of these solutions may require considerable investment that could be difficult to recover. FACTS devices are an alternate solution to address some of those problems. Due to environmental right-of-way and cost problems, there is an interest in better utilization of the existing power system capabilities. The algorithm i.e classical algorithm starts from an initial point to improve the current solution through a orbit which ultimately converges to a current solution. These algorithm include Linear programming (LP) [1], Quadratic programming (QP)[2], Nonlinear programming(NLP)[3] and Mixed integer linear programming(MILP)[4].

In recent decades, an emerging transmission technology called FACTS (flexible AC transmission systems) is extensively employed to increase the power transfer capability of long distance transmission line networks as well as to improve the stability of the transmission systems. FACTS devices can reduce the flows of heavily loaded lines, maintain the bus voltages at desired levels, and improve the stability of the power network [5, 6]. Consequently, they can improve the power system security under contingency situations. The advancement of semiconductor technology launched various FACTS devices which opened up new opportunities for controlling the power flow and extending the load ability of the available power transmission network [7, 8]. It is used in transmission system to reduce the flows in the heavily loaded lines which results in low system loss, improved stability of transmission system by controlling the power flows through the network and increase in load ability [9].

Over the last two decades, many researchers performed a lot of researches on ORPD. Various optimization techniques are evolved to solve ORPD problem. These algorithms are generally divided into two categories, namely, classical mathematical optimization algorithms and intelligent optimization algorithms [10]. The classical algorithms are starting from an initial point, continuously improve the current solution through a certain orbit, and ultimately converging to the optimal solution.

Though, some of these techniques, have a good convergence but most of them suffer from local optimality. Since ORPD is multimodal and non-linear optimization problem and severely depends on the initial guess, the classical techniques are unable to produce global optimal solution.

Various intelligent optimization algorithms known as heuristic techniques are applied to solve ORPD problem. Some of the well popular optimization techniques are evolutionary programming (EP) , genetic algorithm (GA) , simulated annealing (SA) , tabu search (TS) , differential evolution (DE) , particle swarm optimization (PSO) and artificial bee colony (ABC) etc. Recently, a harmony search algorithm (HSA)[8]was developed by Sirjani et al. for simultaneous minimization of total cost, the voltage stability index, voltage profile and power loss of IEEE 57-bus test system using shunt capacitors, SVC and static synchronous compensators (STATCOM). Saravanan et al. presented PSO to find optimal settings and location TCSC, SVC devices for improving system load ability with minimum cost of installation. Optimal reactive power dispatch (ORPD) is an important tool for power system operators for both planning and reliable operation in the present day power systems.

The important aspect of ORPD is to determine the optimal settings of control variables for minimizing transmission loss, improve the voltage profile and voltage stability, while satisfying various equality and inequality constraints [11]. The ORPD problem is in general non-convex and non-linear and exists many local minima. However, in its most general formulation, the OPF

incorporating FACTS is a non-convex and nonlinear optimization problem having both continuous and discrete control variables. The presence of power electronics based FACTS devices and its optimal location problem further complicates the problem solution. Classical calculus-based techniques fail to address these types of problems satisfactorily.

Due to great success of heuristics approaches in solving various non-linear optimization problems, attention has been gradually shifted to applications of population based approaches to handle the complexity involved in nonlinear problem. Recently, many researchers have shown great interest in solving FACTS based OPF problem using evolutionary algorithms such as PSO, GA, evolutionary programming (EP), simulated annealing (SA), ant colony optimization (ACO), differential evolution (DE), biogeography based optimization (BBO), artificial bee colony optimization (ABC) etc [12]. Optimal reactive power dispatch (ORPD) is an important tool for power system operators for both planning and reliable operation in the present day power systems.

The important aspect of ORPD is to determine the optimal settings of control variables for minimizing transmission loss, improve the voltage profile and voltage stability, while satisfying various equality and inequality constraints. The ORPD problem is in general non-convex and non-linear and exists many local minima. Saravanan et al. [13] proposed PSO to find optimal settings and location of TCSC, SVC and UPFC devices with minimum cost of installation for improving system load ability Eslami et al. [14] proposed PSO which is nonlinear time-varying acceleration coefficients (NTVAC-PSO) to damp out of power system oscillations.

In the year of 2011, a new powerful optimization method named chemical reaction optimization (CRO) was presented by Xu et al. [15] to solve nonlinear optimization problems and satisfactory results were obtained in applications due to its enhanced search capability. However, it seems the performance of basic CRO may further be modified to solve non-linear optimization problem. In this article, to improve the performance of CRO. [16] it is integrated with DE for solving nonlinear optimization problems.

DE is a stochastic search algorithm that is originally motivated by the mechanisms of crossover, mutation and natural selection. The potentialities of DE are its simple structure, easy use, local searching property and speediness. Nevertheless, this faster convergence yields in a higher probability of searching toward a local optimum or getting premature convergence. On the other hand, in CRO the control variables are fine-tuned gradually as the process goes on through the operations of on-wall ineffective collision, decomposition, inter-molecular ineffective collision and synthesis. CRO has good exploration ability in finding the region of global minimum.

In order to utilize both the properties of DE and CRO for solution of complex optimization problems, a hybrid technique called hybrid CRO (HCRO) is proposed. The main objective of the present work is to propose HCRO to verify its effectiveness for solving STATCOM based OPF problems. Here the STATCOM is modeled as controllable voltage source ( $E_p'$ ) in series with an impedance [17]. The real part of this admittance represents the copper losses of the coupling transformer and converters, while the imaginary part of this admittance represents the leakage reactance of the coupling transformer. To validate the proposed HCRO-based approach, IEEE 14-bus and IEEE 30-bus systems are considered in this study.

The rest of the paper is as follows: The mathematical formulation of OPF using UPFC controller. A brief description of DE, CRO are demonstrated in section differ evolution and chemical reaction based optimization respectively. HCRO algorithm describes different steps of HCRO technique applied to optimal STATCOM allocation problem which implies the HCRO applied to

STATCOM based OPF problem. Section ‘Simulation results and analysis’ gives numerical results with CRO and DE.

## 2.MATHEMATICAL PROBLEM FORMULATION

### 2.1 Mathematical analysis of Static Synchronous Compensator

Static synchronous compensator (STATCOM) is connected in parallel with the specific bus of a power system. The primary goal of STATCOM is to enhance the reactive power compensation which adjusts the reactive power and voltage magnitude of power system network. It consists of three basic components, namely, transformer, voltage source converter (VSC) and capacitor. The STATCOM is modeled as a controllable voltage source ( $E_p'$ ) in series with an impedance. The real part of this impedance represents the copper losses of the coupling transformer and converter, while the imaginary part of this impedance represents the leakage reactance of the coupling transformer. STATCOM absorbs requisite amount of reactive power from the grid to keep the bus voltage within reasonable range for all power system loading. The injected active and reactive power flow equation of the  $j$ th bus are given below:

$$P_j = G_p' |V_j|^2 - |V_j| \|E_p'\| Y_p' \cos(\delta_k - \delta_j - \theta_p') + \sum_{j=1}^N |V_j| \|V_k\| Y_{jk}' \cos(\delta_k - \delta_j - \theta_{kj}') \quad (1)$$

$$Q_j = -B_p' |V_j|^2 + |V_j| \|E_p'\| Y_p' \sin(\delta_k - \delta_j - \theta_p') - \sum_{j=1}^N |V_j| \|V_k\| Y_{jk}' \sin(\delta_k - \delta_j - \theta_{kj}') \quad (2)$$

The implementation of STATCOM in transmission system introduces two state variables  $|E_p'|, \delta_p'$  however,  $|V_j|$  is known for STATCOM connected bus. It may be assumed that the power consumed by the STATCOM source is zero in steady state.

$$P_{Ep}' = \text{real}(E_p' I_p'^*) \quad (3)$$

$$P_{Ep}' = -G_p' |E_p'|^2 + |E_p'| \|V_j\| Y_p' \cos(\delta_j - \delta_p' + \theta_p') \quad (4)$$

Where  $V_j$  is the voltage at the  $j^{\text{th}}$  bus;  $|Y_p'|$  is the admittance of the STATCOM;  $G_p', B_p'$  are the conductance and susceptance, respectively, of the STATCOM;  $\delta_p'$  is the voltage source angle of the STATCOM;  $E_p'$  is the voltage sources of STATCOM converters.

### 2.2 Objective function

The conventional formulation of ORPD problem determines the optimal setting of control variables such as generator terminal voltages, transformers tap setting, reactive power of shunt compensators, controllable voltage source of STATCOM and its phase angle to minimize the transmission loss while satisfying the operational constraints. However, in order operate the power system in reliable and secure mode, the voltage profile and voltage stability index of the power system are also considered as the objective functions in this study.

#### 2.2.1 Minimization of total real power loss

The objective of transmission loss minimization may be expressed by:

$$f(x, y) = P_{loss} = \sum_{m=1}^{NTL} G_m [V_k^2 + V_j^2 - 2|V_k| \|V_j\| \cos(\delta_k - \delta_j)] \quad (5)$$

where  $f(x, y)$  is the transmission loss minimization objective function;  $P_{loss}$  is the total active power loss;  $G_m$  is the conductance of the  $m^{\text{th}}$  line connected between them  $j^{\text{th}}$  and  $k^{\text{th}}$  buses;  $V_j, V_k$  are

the voltage of the  $j^{th}$  and  $k^{th}$  buses, respectively;  $\partial_j, \partial_k$  are the phase angle of the  $j^{th}$  and  $k^{th}$  bus voltages. X is the vector of dependent variable consisting of load voltages ( $V_{L1}, \dots, V_{LNL}$ ), generators` reactive powers ( $Q_{g1}, \dots, Q_{gNG}$ ), transmission lines` loadings ( $S_{L1}, \dots, S_{LNTL}$ ), controllable voltage source of STATCOM ( $E_{p1}, \dots, E_{pPN}$ ) and phase angle of STATCOM ( $\partial_{p1}, \dots, \partial_{pPN}$ ); y is the vector of independent variables consisting of generators` voltages ( $V_{g1}, \dots, V_{gNG}$ ), transformers` tap settings ( $T_1, \dots, T_{NT}$ ), reactive power injections ( $Q_{i1}, \dots, Q_{iNC}$ ) and voltage of the buses where STATCOMs are used ( $V_{STATCOM1}, \dots, V_{STATCOMn}$ ). Therefore, the independent and dependent vectors may be expressed as:

$$x = [V_h, \dots, V_{LNL}, Q_{g1}, \dots, Q_{gNG}, S_h, \dots, S_{LNTL}, E_{p1}, \dots, E_{pPN}, \partial_{p1}, \dots, \partial_{pPN}] \quad (6)$$

$$y = [V_{g1}, \dots, V_{gNG}, T_1, \dots, T_{NT}, Q_i, \dots, Q_{iNC}, V_{STATCOM1}, \dots, V_{STATCOMn}] \quad (7)$$

Where NG; NL are the number of generator and load buses; NTL; NT; NC are the number of transmission lines, regulating transformers and shunt compensators, respectively.

### 2.2.2 Minimization of voltage deviation

Since bus voltage is one of the most important security and service quality indexes of the power system, the minimization of deviations of voltages from desired values is considered as another objective in this study. The objective function of voltage profile improvement, i.e. voltage deviation minimization at load buses, may be expressed as:

$$\begin{cases} V_{Lq}, V_{Lq}^{Ss}, NL, f_{2(x,y)} \\ f_{2(x,y)} = \min(\sum_{q=1}^{NL} |V_{Lq} - V_{Lq}^{Ss}|) \end{cases} \quad (8)$$

where  $V_{Lq}$  is the voltage at the  $j^{th}$  load bus;  $V_{Lq}^{Ss}$  is the desired voltage at the  $j^{th}$  load bus, usually set to 1.0 p.u.

### 2.2.3 Minimization of L-index

It is very important to maintain constantly acceptable bus voltage at each bus under normal operating conditions. However, when the system is subjected to a disturbance, the system configuration is changed. The non-optimized control variables may lead to progressive and uncontrollable drop in voltage resulting in an eventual widespread voltage collapse. In this work, voltage stability enhancement is achieved through minimizing the voltage stability indicator L-index. The indicator values vary in the range between 0 and 1. The L-index of a power system is briefly discussed below: For a multi-node system, the relation among voltage and current of load and generator buses may be expressed as follows:

$$\begin{bmatrix} I_l \\ I_g \end{bmatrix} = \begin{bmatrix} y_{li} & y_{lg} \\ y_{gi} & y_{gg} \end{bmatrix} \begin{bmatrix} V_l \\ V_g \end{bmatrix} \quad (9)$$

By matrix inversion, the above equation may be rearranged as follows:

$$\begin{bmatrix} V_l \\ I_g \end{bmatrix} = \begin{bmatrix} Z_{li} & F_{lg} \\ K_{gi} & Y_{gg} \end{bmatrix} \begin{bmatrix} I_l \\ V_g \end{bmatrix} \quad (10)$$

The sub-matrix  $F_{lg}$  may be expressed as under:



$$F_{i'g'} = -[y_{11}]^{-1}[y_{i'g'}] \quad (11)$$

The voltage stability index of the  $K^{th}$  bus may be expressed by

$$L_k = |1 - \sum_{j=1}^{N_g} F_{kj} \frac{V_j}{V_k}| \quad k=1,2,\dots,N_l \quad (12)$$

Where  $V_{g'}$ ;  $V_{l'}$  are the vectors of the bus voltage of the generator and load buses, respectively;

$I_{g'}$ ;  $I_{l'}$  are the vectors of the bus currents of the generator and load buses, respectively.  $Z_{i'}$ ,  $F_{i'g'}$ ,  $K_{g'i'}$ ,  $Y_{g'g'}$  are the sub-matrices obtained by partial inversion of the admittance matrix,  $N_{g'}$ ;  $N_{l'}$  are the number of generator and load buses, respectively.

To move the system far away from the voltage collapse point, the voltage stability index needs to minimize. The global L-index indicator of the power system is expressed as follows:

$$L_{\max} = \max(L_1, L_2, \dots, L_{N_l}) \quad (13)$$

Therefore, to enhance the voltage stability and to move the system far from the voltage collapse margin, the objective function may be represented as follows:

$$f_{3(x,y)} = \min L_{\max} \quad (14)$$

### 2.3 Constraints

The ORPD incorporating STATCOM is subjected to the following constraints:

#### i. Equality constraints

The equality constraints which describe the load flow equations are represented by

$$\begin{cases} \sum_{j=1}^{NB} P_{gj} - P_{lj} = \sum_{j=1}^{NB} \sum_{k=1}^{NB} V_j V_k |g_{jk} \cos \theta_{jk} + b_{jk} \sin \theta_{jk}| \\ \sum_{j=1}^{NB} q_{gj} - q_{lj} = \sum_{j=1}^{NB} \sum_{k=1}^{NB} V_j V_k |g_{jk} \sin \theta_{jk} - b_{jk} \cos \theta_{jk}| \end{cases} \quad (15)$$

Where  $p_{gj}$ ,  $p_{lj}$  are the active power generation and demand, respectively, of the  $j$ th bus;  $q_{gj}$ ,  $q_{lj}$  are the reactive power generation and demand, respectively, of the  $j$ th bus;  $g_{jk}$ ,  $b_{jk}$  are the conductance and susceptance, respectively, of the line connected between them  $j^{th}$  bus and  $k^{th}$  bus and NB is number of buses.

#### 2.3.2 Inequality constraints:

##### (i) Generator constraints:

$$\begin{cases} V_{gj}^{\min} \leq V_{gj} \leq V_{gj}^{\max} \\ P_{gj}^{\min} \leq P_{gj} \leq P_{gj}^{\max} \\ q_{gj}^{\min} \leq q_{gj} \leq q_{gj}^{\max} \end{cases} \quad (16)$$

##### (ii) Load bus constraints:

$$V_{lj}^{\min} \leq V_{lj} \leq V_{lj}^{\max} \quad ; j=1, 2, \dots, NL \quad (17)$$

##### (iii) Transmission line constraints:

$$S_{lj} \leq S_{lj}^{\max} \quad j=1,2,\dots,NL \quad (18)$$

(iv) Transformer tap constraints:

$$t_j^{\min} \leq t_j \leq t_j^{\max} \quad ; j=1,2,\dots,\text{NT} \quad (19)$$

(v) Shunt compensator constraints:

$$q_{cj}^{\min} \leq q_{cj} \leq q_{cj}^{\max} \quad ; j=1,2,\dots,\text{NC} \quad (20)$$

(vi) UPFC series source constraints:

$$E_{pj}^{\min} \leq E_{pj} \leq E_{pj}^{\max} \quad ; j=1,2,\dots,\text{NUPFC} \quad (21)$$

$$\partial_{pj}^{\min} \leq \partial_{pj} \leq \partial_{pj}^{\max} \quad ; j=1,2,\dots,\text{NUPFC} \quad (22)$$

Where  $V_{gj}^{\min}, V_{gj}^{\max}$  are the voltage operating limits of the  $j^{\text{th}}$  generator bus;  $P_{gj}^{\min}, P_{gj}^{\max}$  are the active power generation limits of the  $j^{\text{th}}$  bus;  $Q_{gj}^{\min}, Q_{gj}^{\max}$  are the reactive power generation limits of the  $i^{\text{th}}$  bus;  $V_{lj}^{\min}, V_{lj}^{\max}$  are the voltage limits of the  $j^{\text{th}}$  load bus;  $S_{lj}, S_{lj}^{\max}$  are the apparent power flow and maximum apparent power flow limit of the  $j^{\text{th}}$  branch;  $t_j^{\min}, t_j^{\max}$  are the tap setting limits of the  $j^{\text{th}}$  regulating transformer;  $q_{cj}^{\min}, q_{cj}^{\max}$  are the reactive power injection limits of the  $j^{\text{th}}$  shunt compensator ;  $\partial_{pj}^{\min}, \partial_{pj}^{\max}$  are the phase angle limits of the  $j^{\text{th}}$  STATCOM;  $E_{pj}^{\min}, E_{pj}^{\max}$  are the voltage limits of the  $j^{\text{th}}$

STATCOM; NG;NL;NTL;NT;NC are the number of generator bus, load bus, transmission line, regulating transformer and shunt compensator respectively.

### 3.ALGORITHMS

The performance of ORPD problem is highly susceptible to the selection of its control variables. Thus to extract optimal performance of ORPD, proper tuning of its control variables is essential. This work proposes some popular stochastic optimization techniques namely differential evolution (DE), chemical reaction optimization (CRO), and hybrid CRO to ameliorate performance of the concerned power systems. The aforesaid algorithms are briefly discussed below:

#### 3.1 ` Differential Evolution

Differential Evolution (DE) is a population based optimization technique proposed by Storm and Price in 1995. It is also a population based anti-heuristic algorithm and is capable of handling non-differential, non-linear and multi model objective function. It consists of two populations

- (a) New/Child population
- (b) Old/Parent population

It is highly sensitive to the definition of input parameters. It is implemented by using 3 generic operators namely, Mutation, Crossover and Selection. The algorithm steps of DE are as follows:

##### (i) Initialization

The population is initiated by using randomly generated individual within operating constraints as

$$X_{ij} = X_j^{\min} + rand^*(X_j^{\max} - X_j^{\min}) \quad \text{Where } \begin{matrix} i = 1,2,\dots, NP \\ j = 1,2,\dots, ND \end{matrix} \quad (23)$$

Where,  $X_{ij}$  initialized by  $j^{\text{th}}$  control variable of  $i^{\text{th}}$  population set.  
 $X_j^{\min}, X_j^{\max}$  minimum and maximum limits of  $j^{\text{th}}$  control variables.  
 $rand^*$  uniformly generated random variables between [0,1 ].  
 $N_D$  number of control variables.

$N_p$  population size.

**(ii) Mutation**

It plays an important role in differential evolution. Its main objective is to produce offspring on target vector. Each target vector undergoes mutation to produce a mutant vector which is obtained by adding the weighted difference between two randomly chosen mutant vectors to the third one.

Mutant vector is generated by:

$$V_{i,G+1} = X_{r1G} + F^* (X_{r2G} - X_{r3G}) \quad (24)$$

Where  $V_{i,G+1}$  mutant vector at  $i^{th}$  population &  $G+1$  iteration.

$X_{r1G}, X_{r2G}, X_{r3G}$  randomly chosen vectors.

$F^*$  input variable or measure between 2 weighted variables between  $[0, 2]$  where  $r_1 \neq r_2 \neq r_3 \neq i$ .

**(iii) Crossover**

It is implemented to increase the diversity of population. Crossover is employed to generate a trial vector which is obtained by mixing mutant vector to target vector. If trial vector produces better solution than target vector then trial vector is generated by:

$$\begin{cases} U_{i,G+1} = V_{i,G+1} & \text{if } rand \leq CR \\ X_{i,G+1} = U_{i,G+1} & \text{if } rand \geq CR \end{cases} \quad CR \in 0.5,1 \quad (25)$$

where  $U_{i,G+1}$  is the trial Vector,  $V_{i,G+1}$  is the mutant Vector,  $X_{i,G+1}$  is the target Vector

**(iv) Selection**

It is the procedure of producing better offspring from target or trial vector .

$$\begin{cases} X_{i,G+1} = U_{i,G+1} & \text{if } U_{i,G+1} \leq X_{i,G+1} \\ = X_{i,G} & \text{otherwise} \end{cases} \quad (26)$$

**3.2 Chemical reaction based optimization**

Chemical reaction optimization (CRO) was introduced by Lam and Li in the year 2010. It is a new optimization technique based on the various chemical reactions occur among the molecules. A molecule consists of several atoms and is characterized by the atom type, bond length and torsion any change in the atom type makes the molecules different from others. Each molecule has two kinds of energies PE (potential energy) and KE (kinetic energy). PE represents the objective function of a molecule while the KE of a molecule represents its ability of escaping from a local minimum. During the CRO process, the following four type of elementary reactions are likely to happen. These are on wall ineffective collision, decomposition, inter-molecular ineffective collision and synthesis. These reactions can be categorized into single molecular reactions and multiple molecular reactions. The on-wall ineffective collision and decomposition reactions are single molecular reactions, while the intermolecular ineffective collision and synthesis reactions are of the latter category.

**(i) On-wall ineffective collision:**

In this reaction process each molecule hits the wall of the container and generates a new molecule whose molecular structure is closed to the original one. Since, the On-wall ineffective collision is not so severe, the resultant molecular structure is not too different from the original one. A molecule ‘ms’ collides into the wall is allowed to change to another molecule ‘ms1’, if the constraint described below is satisfied.

$$PE^{ms'} + KE^{ms'} \geq PE^{ms_i} \quad (27)$$

**(ii) Decomposition:**

A single compound breaks down into two or more molecules in the decomposition process. In this reaction, the newly formed molecules are far away from the original molecule. As compared with on-wall ineffective collision, the generated molecules have greater change in the potential energy than the original ones. The molecule  $m$ , hits a wall of the container and participate in decomposition reaction, to generate two molecules  $ms_1$  and  $ms_2$  if the following inequality constraint holds

$$PE^{ms} + KE^{ms} \geq PE^{ms_1} + PE^{ms_2} \quad (28)$$

**(iii) Intermolecular ineffective collision:**

This chemical reaction takes place when two different molecules react among themselves, forming two different molecules. However, in this reaction, the molecular structures of the newly generated molecules are closed to the original molecules. Therefore in this collision, the molecules react much less vigorously than decomposition collision. When two molecules, ' $m^1$ ' and ' $m^2$ ', collide with each other, they may form to two new molecules, ' $m_1^1$ ' and ' $m_2^1$ ', if the following inequality holds

$$KE^{ms_1} + PE^{ms_1} + KE^{ms_2} + PE^{ms_2} \geq PE^{ms_1} + PE^{ms_2} \quad (29)$$

**(iv) Synthesis:**

The synthesis reaction is opposite to the decomposition reaction. In this reaction two or more reactants combine together to form an entirely different new molecule. Synthesis collision allows the molecular structure to change in a larger extent. The two molecules ' $m_1^1$ ' and ' $m_2^1$ ' collide with each other and form a new molecule  $m$  if the following condition is satisfied.

$$KE^{ms_1} + PE^{ms_1} + KE^{ms_2} + PE^{ms_2} \geq PE^{ms} \quad (30)$$

The kinetic energy for the newly formed molecule ' $m$ ' is modified as follows:

$$KE^{ms_1} + PE^{ms_1} + KE^{ms_2} + PE^{ms_2} - PE^{ms} = KE^{ms} \quad (31)$$

The various steps for implementing the CRO algorithm can be summarized as follows:

- Step 1: The various input parameters of the CRO algorithm are initialized. The molecular structures of the molecules are generated randomly. The molecular structures of the molecules represent various feasible solution vectors.
- Step 2: The value of the objective function of the individual feasible solution set represents the potential energy (PE) of the individual molecule. An initial kinetic energy (KE) is assigned to all the molecules.
- Step 3: Depending upon the PE values, sort the population and in order to retain the best solutions intact, few best molecules are kept as elite molecules.
- Step 4: Collision operations are performed on non-elite molecules. In this process, one molecule  $ms$  is selected randomly from the population and one molecule  $ms_1$  is generated using mutation operation as described below. To allow the algorithm to escape from a local minimum, the on-wall ineffective collision operations.

$$ms_2^{i,j} = ms_1^{k,j} + F * (ms_1^{m,j} - ms_1^{n,j}), i=1, \dots, N_p \quad (32)$$

$ms_1^{k,j}$ ;  $ms_1^{m,j}$  and  $ms_1^{n,j}$  are the  $j$ th components of three different molecules chosen randomly from the current population. If there is enough energy for the new molecule to be generated, i.e. if criterion (33) is satisfied, replace the original molecule with the new one, and update the relevant KE using (34).

$$KE^{m's} + PE^{m's} \geq PE^{ms_1} \quad (33)$$

$$KE^{m's_1} = rand * (KE^{m's} + PE^{m's} - PE^{m's_2}) \quad (34)$$

Step 5: For each decomposition operation, two molecules are selected from the population and two molecules are generated by performing the crossover operation of DE. Afterward, they are tested against the synthesis criterion:

$$KE^{m's} + PE^{m's} \geq PE^{m's} + PE^{m's_1} \quad (35)$$

If this criterion is satisfied by the selected molecules, replace the original molecules by the newly generated molecules and update the KE of the new molecules using (36) and (37).

$$KE^{m's_1} = rand * [KE^{m's} + PE^{m's} - (PE^{m's_1} + PE^{m's_1})] \quad (36)$$

$$KE^{m's_1} = (1 - rand) * [KE^{m's} + PE^{m's} - (PE^{m's_1} + PE^{m's_1})] \quad (37)$$

Step 6: To enhance the search space, the inter-molecular ineffective collision is applied on each molecule to update its molecular structure. The intermolecular ineffective collision occurs when two molecules collide and then produce two new molecules. To perform this reaction, two molecules  $m's_1''$  and  $m's_2''$  from the population are selected and two new molecules  $m's_1''$  and  $m's_2''$  are generated by performing the crossover operation of DE. The original molecules  $m's_1'$  and  $m's_2'$  are replaced by the new molecules  $m's_1''$  and  $m's_2''$  if the newly generated molecules have better fitness value (PEv). The KE of the molecules  $m's_1''$  and  $m's_2''$  are modified using (38) and (39)

$$KE^{m's_1''} = rand * [PE^{m's_1'} + KE^{m's_1'} + PE^{m's_2'} + KE^{m's_2'} - (PE^{m's_1''} + PE^{m's_2''})] \quad (38)$$

$$KE^{m's_2''} = (1 - rand) * [PE^{m's_1'} + KE^{m's_1'} + PE^{m's_2'} + KE^{m's_2'} - (PE^{m's_1''} + PE^{m's_2''})] \quad (39)$$

Step 7: Lastly, the molecules participate in synthesis collision operation to update their molecular structure. Two molecules  $ms_1$  and  $ms_2$  are selected randomly from the population set and one molecule  $ms_0$  is generated by performing the crossover operation.

If the newly generated molecule gives better function value (PE), the new molecule is included and the original molecules are excluded. The new molecule becomes

$$KE^{m's_1''} = rand * (KE^{m's_1'} + PE^{m's_1'} + KE^{m's_2'} + PE^{m's_2'} - PE^{m's_1''}) \quad (40)$$

Step 8: The feasibility of each solution is checked by satisfying its operational constraints.

Step 9: Sort the solutions from best to worst and replace the worst solution by the best elite solutions.

### 3.3 HCRO algorithm

The CRO algorithm that emphasizes on exploring the entire search space and a local version of DE algorithm that emphasizes on exploiting the local search space are combined together to make an impact on the performance of the algorithm in terms of the solution quality and convergence speed. Such an algorithm is the HCRO algorithm.

The general steps of the HCRO algorithm is summarized as follows:

Step 1: Initialize several numbers of molecules by randomly generated molecular structure depending upon the population size. The molecular structure of each molecule represents a potential solution to the given problem.

Step 2: The fitness values of the specific problem of the population are assigned as the potential energy (PE) of the individual molecule. A random kinetic energy (KE) is set to the different molecules.

Step 3: Based on the PE values (fitness values) best solutions are retained by elite molecules.

Step 4: The non-elite molecules are modify using on-wall ineffective collision operations as described below:

Step 4.1: One molecule  $ms$  is selected randomly.

Step 4.2: Using the mutation operation of DE (described in Section ‘Differential evolution’) new molecule generated which may mathematically  $ms'_1$  is be expressed as:

$$ms'_1 = ms'_k + f_m^* (ms'_m - ms'_n) \quad (41)$$

Condition  $E_k^{ms'} + E_p^{ms'} \geq E_p^{ms'_1}$  is satisfied and the KE of the molecule  $ms'_1$  evaluated by

$$E_k^{ms'_1} = rand(0,1)^* (E_k^{ms'} + E_p^{ms'} - E_p^{ms'_1}) \quad (42)$$

Step 5: Decomposition operation is performed to modify the molecular structure of the molecules by the following steps:

Step 5.1: Two molecules, one molecule  $ms$  from the population and another randomly generated molecule  $ms'_1$  are selected for decomposition operation.

Step 5.2: Crossover operation of DE is applied on  $ms'$  and  $ms'_1$  to generate two new molecules  $ms'_1$  and  $ms'_2$ .

Step 5.3: Potential energy  $E_p^{ms}$  and  $E_p^{ms'_1}$  of molecules  $ms'$  and  $ms'_1$  are evaluated. I

$E_k^{ms'} + E_p^{ms'} \geq E_p^{ms''} + E_p^{ms'_1}$  molecule  $ms$  is deleted and the molecules  $ms_0$  and  $ms_01$  are pushed into the population. Modify the KE of  $ms'$  and  $ms'_1$  as below:

$$E_k^{ms''} = rand(0,1)^* [E_k^{ms'} + E_p^{ms'} - (E_p^{ms''} + E_p^{ms'_1})] \quad (43)$$

$$E_k^{ms'_1} = [1 - rand(0,1)]^* [E_k^{ms'} + E_p^{ms'} - (E_p^{ms''} + E_p^{ms'_1})] \quad (44)$$

Step 6: Inter-molecular ineffective collisions are made to modify the molecular structure each molecule using the following steps:

Step 6.1: Randomly select two molecules  $ms_1$  and  $ms_2$  from the population.

Step 6.2: Two new molecules  $ms_1^1$  and  $ms_2^1$  are generated using crossover operation of DE.

Step 6.3: Evaluate potential energy  $E_p^{ms_1^1}$  and  $E_p^{ms_2^1}$  of molecules  $ms_1^1$  and  $ms_2^1$  replace molecules

by molecules  $ms_1^1$  and  $ms_2^1$ , respectively, if

$$E_p^{ms_1^1} + E_k^{ms_1^1} + E_p^{ms_2^1} + E_k^{ms_2^1} \geq E_p^{ms_1^1} + E_p^{ms_2^1} \quad (45)$$

The KE of the molecules  $ms^1$  and  $ms^2$  are evaluated using

$$E_k^{ms_1^1} = rand(0,1)^* [E_p^{ms_1^1} + E_k^{ms_1^1} + E_k^{ms_2^1} - (E_p^{ms_1^1} + E_p^{ms_2^1})] \quad (46)$$

$$E_k^{ms_2^1} = [1 - rand(0,1)]^* [E_p^{ms_1^1} + E_k^{ms_1^1} + E_p^{ms_2^1} + E_k^{ms_2^1} - (E_p^{ms_1^1} + E_p^{ms_2^1})]$$

(47)

Step 7: Synthesis collision operation are performed to update the molecular structure of the molecules using the following steps:

Step 7.1: Two molecules  $ms^1$  and  $ms^2$  are randomly selected from the population set.

Step 7.2: Apply the conventional cross over operation of GA on  $ms^1$  and  $ms^2$  by considering them as parents chromosomes and generate a child chromosome  $ms_1^1$ .

Step 7.3: Evaluate  $E_p^{ms_1^1}$  of the molecule  $ms_1^1$ . The molecules  $ms^1$  and  $ms^2$  are omitted and the molecule  $ms_1^1$  is push into the population if

$$E_k^{ms^1} + E_p^{ms^1} + E_k^{ms^2} + E_p^{ms^2} \geq E_p^{ms_1^1} \quad (48)$$

The KE of the new molecule  $ms_1^1$  is calculated using

$$E_k^{ms_i'} = rand(0,1) * (E_k^{ms_i} + E_p^{ms_i} + E_k^{ms_2} + E_p^{ms_2} - E_p^{ms_i'}) \quad (49)$$

Step 8: The feasibility of a problem solution is verified and the infeasible solutions are replaced by feasible solution set.

Step 9: The updated molecules are sorted.

Step 10: The best elite molecules are replaced by the worst molecule.

Step 11: The processes of generating new molecules and selecting those with better function values are continued until the given stopping conditions are satisfied. The iteration process can be stopped after a fixed number of generations or when any significant improve improvement in the solution does not occur. In this paper, CRO and HCRO are run for a fixed number of generations.

## 4. SIMULATION RESULTS AND DISCUSSION

To enhance the efficient of the proposed HCRO approach, two case studies are used in this paper (IEEE-30 bus and IEEE-57 bus systems) of ORPD problems are used in the simulation study. The programs which are present all are written in MATLAB7.0 and run on a PC with core i3 processor, 2.50GHZ, 4GB RAM. The results of the ORPD problem obtained by HCRO are compared with those obtained by HCRO are compared with those obtained by CRO,DE and other techniques such as canonical GA(CGA)[26], the adaptive inertia weight(PSO-W)[26], PSO with a constriction factor(PSO-cf)[26] the comprehensive learning particle swarm optimization(CLPSO)[26], the standard version of PSO(PSO)[27],local search DE with self-adapting control parameters(L-SACP-DE)[26], seeker optimization algorithm(SOA) [26],teaching based optimization(TLBO)[18], quasi-oppositional TLBO(QOTLBO)[18], strength pare to evolutionary algorithm(SPEA)[19], GA-1[20] and GA-2[21] multi-objective PSO(MOPSO-1)[22], DE-1[23], oppositional GSA(OGSA)[24], multi-objective PSO(MOPSO-2)[25] multi-objective improved PSO(MOIPSO)[25], multi-objective chaotic improved PSO(MOCIPSO)[25] available in the literature. It should be carefully chosen as the performance of any algorithm depends on its input parameters. Table-1 gives the results of comparison of simulation obtained by different algorithm without using STATCOM devices.

### 4.1 IEEE 30- bus system

IEEE 30-bus system is used at first to evaluate the correctness and relative performance of the proposed HCRO method. This system consists of 6 generators, 9 shunt compensators, 4 regulating transformers and 41 transmission lines.

The voltage limits i.e. the maximum and minimum voltage limits at all the buses are taken as 1.10p.u and 0.9p.u, respectively. The resistance and reactance of STATCOM converters is 0.01p.u and 0.1p.u. The limits of phase angle of STATCOM areas  $-20^\circ \leq \delta_p' \leq 0^\circ$ . Also the upper and lower tap setting of regulating transformers are 1.10p.u and 0.9p.u, respectively. The performance of the proposed HCRO method is explained by applying ORPD problem without STATCOM (case 1) and ORPD with STATCOM (case 2) and its results are compared with those of other methods.

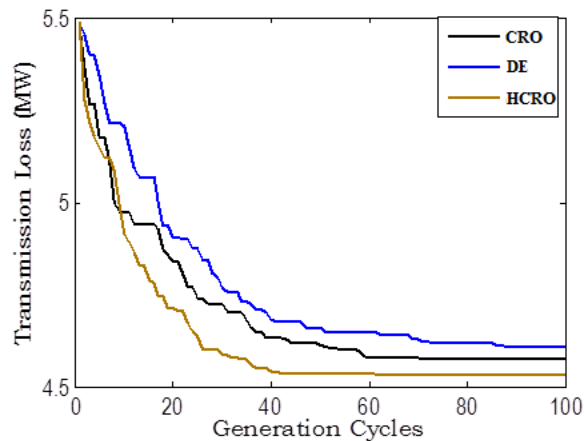
#### 4.1.1 Case A: Transmission loss minimization

##### (i) ORPD without STATCOM device

The Symphony of the proposed HCRO method along with DE and CRO is initially verified by applying it to minimize transmission loss of IEEE 30-bus system without STATCOM.

The transmission loss and the optimal settings of control variables obtained from HCRO, DE and CRO algorithm are given in table1. The results show that the transmission loss found by the proposed

HCRO method is lower than DE and CRO. Figure 1 shows the real power loss variation against the number of iterations for the HCRO, DE and CRO algorithm. However, 50 trials with different



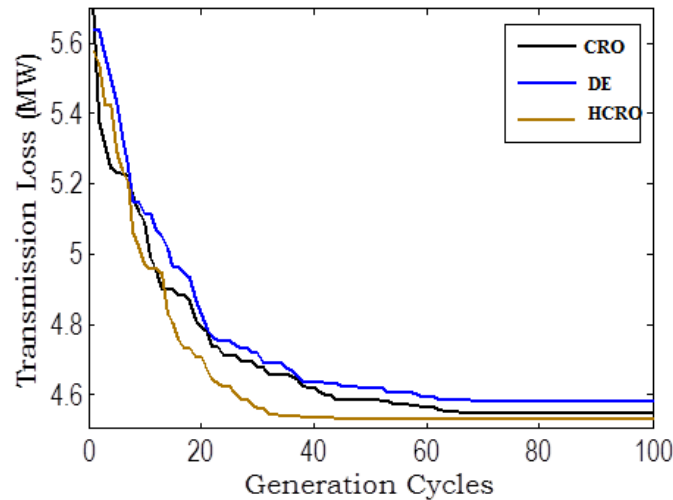
**Fig.1** Convergence characteristics of different algorithms for transmission loss without STATCOM (IEEE 30-bus system)

Populations are carried out to test the robustness of the HCRO algorithm and its statistical results are compared with those of TLBO [18], QOTLBO [18], SPEA [19], and GA-1 [20] and GA-2 [21]. The statistical results are obtained in Table-2 which gives the best, worst and the average results obtained by HCRO are more or less same and variation between them is negligible. These facts describe strongly about the robustness of the proposed HCRO for the ORPD problem. The worst and mean less of SPEA, GA-1, GA-2 are not available (NA) in the literature.

**(ii) ORPD with STATCOM**

To solve ORPD with STATCOM of the same test system it need to be applied, in order to the check the feasibility of the proposed method for this complicated work. The results for simulation of optimal position of STATCOM, simultaneously results of transmission loss, the controlled variables and its voltage rating are obtained by HCRO, DE and CRO are shown in Table-3. Also, the results indicate that the proposed HCRO algorithm gives more reduction loss (4.4812MW) as compared to DE (4.5493MW) and CRO (4.5297MW). The convergence of transmission loss which gives the minimal value with evolution generation are shown in figure 2 which certifies the results of HCRO, DE and CRO are reported Table-4. From Table-4 it is evident that HCRO not only has found the highest quality results among the all algorithm compared but possesses the highest probability of finding the better solution for the problem under consideration.





**Fig. 2** Convergence characteristics of different algorithms for transmission loss with STATCOM (IEEE 30-bus system)

### **.1.3 Case B: Voltage deviation minimization**

The results which are obtained for this objective for this objective function by HCRO, DE and CRO without and with STATCOM are given in Table-1 and Table-3 respectively. From the simulation results it is observed that the voltage deviation is improved by incorporating FACTs device i.e STATCOM by HCRO, DE and CRO method. Also it is viewed that voltage deviation using proposed HCRO method is better as compared to that obtained by DE and CRO algorithms. The statistical results for voltage deviation minimization objective illustrated in Table-3 and Table 4. Show the superiority of the proposed CRO method over other approaches.

#### **4.1.3 Case C: Minimization of L-index voltage stability**

In order to investigate further the efficiency of the proposed HCRO method, it is applied on the same IEEE 30-bus system to minimize voltage stability index. Tables3 and 4 show the optimal settings of control variables, optimal locations and optimal parameter settings of STATCOM obtained by applying HCRO, DE and CRO techniques for normal and FACTs based ORPD problem. For voltage stability index minimization objective, before using FACTs devices in the transmission network. However, the best L-index is obtained using HCRO method for both the cases (i.e. with and without STATCOM).

## **4.2 IEEE 57-bus system**

In order to assess the robustness and effectiveness of the proposed HCRO method, a larger test system consisting of 57buses order to asses with and without STATCOM is considered to solve ORPD problem. This system (IEEE 57-bus) consists of seven generator buses(the bus 1 is the slack bus and buses 2,3,6,8,9 and 12 are PV buses), fifty load buses and 80 branches, in which branches(4-12, 20-21, 24-26, 7-29, 32-34, 11-41, 15-45, 14-46, 10-51, 13-49, 11-43, 40-56, 39-57, and 9-55) are tap changing transformers. In addition, buses 8, 25 and 53 are selected as shunt compensated buses. In this study, the allowed steps for tap changers are between 0.9 and 1.1p.u the allowed voltage changes are between 0.95 and 1.05. There by certain simulations are carried out for conventional ORPF problem and STATCOM based ORPD problems to test and validate the robustness of the proposed algorithm.

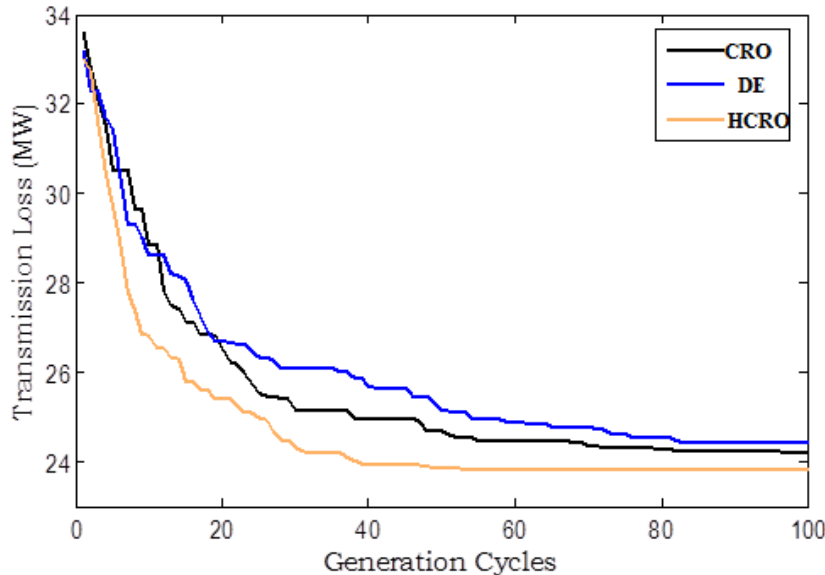
### **4.2.1 Case A: Transmission loss minimization**

**(i) ORPD with STATCOM device**

By HCRO, DE and CRO method the optimal settings of control variables are given in Table-5. It is observed that all the control variables and state variables are in their specified limits. The results of this comparison are given in Table-6. All the comparison is carried out with the same control variable limits.

**(ii) ORPD with STATCOM device**

The symphony of the HCRO method is evaluated further by implementing proposed method on IEEE57-bus system in order to minimize the transmission loss STATCOM which is based on power system network. The simulation results of HCRO, DE and CRO are given in Table-7. It can be observed from the results, that the proposed HCRO algorithm gives the best performance as compared with DE and CRO methods. Also to verify the robustness, the DE, CRO and HCRO algorithms are executed 50 trials with different starting points. It is observed and worth mentioning that the best, worst and average loss obtained by the HCRO method gives the better results than those obtained by DE and CRO methods. The convergence of optimal solution using HCRO, DE and CRO is shown in figure 3.



**Fig.3** Convergence characteristics of different algorithms for transmission loss with STATCOM (IEEE 57-bus system)

#### 4.2.2 Case B: Voltage deviation minimization

The corresponding results obtained by the different methods are listed in Tables 5 and 7. The voltage deviation value obtained by DE, CRO and HCRO method is 0.6919 p.u, 0.6724 p.u and 0.6700 p.u, respectively, for ORPD without FACTS. Here DE, CRO and HCRO approaches are applied on the same test system with the objective of voltage deviation minimization with and without STATCOM devices. After incorporating the STATCOM, voltage deviation value obtained by DE, CRO and HCRO method is 0.6803 p.u, 0.6533 p.u, 0.6523 p.u, respectively. By incorporating STATCOM in optimal location, voltage deviation has been reduced significantly. Thus the simulation results obtained, indicate that reduction of voltage deviation by HCRO is best among all the discussed algorithms for both ORPD which is normal one and STATCOM based ORPD problems.

#### 4.2.3 Case C: Minimization of L-index voltage stability

The optimal control variables TL, SVD and L-index values obtained using DE, CRO and HCRO approaches in the IEEE 57- bus power system for L-index minimization objective of normal ORPD and STATCOM based ORPD are described elaborately in the Tables 5 and 7 respectively. The L-

index reduction accomplished using the HCRO approach which is better than that obtained by the other approaches. Thus, the conclusion can be drawn that HCRO is better than all other algorithms. Also the control variables optimized by the various methods are acceptably kept within the limits. Fig. 4 shows that the convergence performance of algorithms with the evolution process. Thus the statistical results of L-index minimization objective for normal ORPD and STATCOM based ORPD problems are illustrated in Tables 6 and 8 respectively.

**List of Tables**

**Table 1**  
Comparison of simulation results obtained by different algorithms without STATCOM

Real power loss minimization			Voltage deviation minimization			Voltage stability index minimization			
Algorithms	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]
V1 (p.u.)	1.1000	1.0972	1.0998	1.0081	1.0089	1.0092	1.0987	1.0972	1.0998
V2 (p.u.)	1.0941	1.0869	1.0939	1.0031	1.0044	1.0050	1.0968	1.0869	1.0939
V5 (p.u.)	1.0746	1.0640	1.0743	1.0198	1.0218	1.0195	1.0991	1.0640	1.0743
V8 (p.u.)	1.0764	1.0686	1.0762	1.0039	1.0041	1.0031	1.0990	1.0686	1.0762
V11 (p.u.)	1.0998	1.0990	1.0997	1.0278	1.0027	1.0390	1.0982	1.0990	1.0997
V13 (p.u.)	1.1000	1.0981	1.0999	1.0264	1.0284	1.0144	0.9506	1.0981	1.0999
TC6-9	1.0387	1.0290	1.0380	1.0420	1.0142	1.0551	0.9009	1.0290	1.0380
TC6-10	0.9000	0.9034	0.9011	0.9021	0.9004	0.9001	0.9030	0.9034	0.9011
TC4-12	0.9726	0.9730	0.9727	1.0076	1.0136	0.9937	1.0998	0.9730	0.9727
TC27-28	0.9628	0.9612	0.9636	0.9655	0.9667	0.9663	0.9012	0.9612	0.9636
QC10 (p.u.)	0.0500	0.0497	0.0499	0.0500	0.0500	0.0499	0.0500	0.0497	0.0499
QC12 (p.u.)	0.0500	0.0484	0.0499	0.0106	0.0199	0.0424	0.0004	0.0484	0.0499
QC15 (p.u.)	0.0500	0.0495	0.0499	0.0500	0.0498	0.0500	0.0001	0.0495	0.0499
QC17 (p.u.)	0.0500	0.0466	0.0499	0.0000	0.0000	0.0000	0.0500	0.0466	0.0499
QC20 (p.u.)	0.0420	0.0423	0.0422	0.0500	0.0500	0.0500	0.0497	0.0423	0.0422
QC21 (p.u.)	0.0500	0.0490	0.0499	0.0500	0.0499	0.0500	0.0496	0.0490	0.0499
QC23 (p.u.)	0.0262	0.0255	0.0263	0.0500	0.0500	0.0500	0.0499	0.0255	0.0263
QC24 (p.u.)	0.0500	0.0493	0.0500	0.0500	0.0500	0.0500	0.0487	0.0493	0.0500
QC29 (p.u.)	0.0224	0.0277	0.0228	0.0248	0.0497	0.0258	0.0493	0.0277	0.0228
SVD (p.u.)	2.0973	1.9878	2.0888	0.0841	0.1029	0.0849	1.8791	1.9878	2.0888
TL (MW)	4.5304	4.5749	4.5322	5.8637	5.8973	5.8067	7.9377	4.5749	4.5322
L- index	0.1263	0.1321	0.1298	0.1356	0.1423	0.1411	0.1148	0.1302	0.1156

**Table 2**  
Statistical comparison (50 trials) among various algorithms for IEEE 30-bus without STATCOM

Real power loss								
Techniques	TLBO[18]	QOTLBO[18]	SPEA[19]	GA-1[20]	GA-2[21]	CRO	DE	HCRO
Best loss (MW)	4.5629	4.5594	5.1170	4.5800	4.5550	4.5322	4.5749	4.5304
Mean loss (MW)	4.5695	4.5601	NA	NA	NA	4.5413	4.6414	4.5391
Worst loss (MW)	4.5748	4.5617	NA	NA	NA	4.5476	4.7328	4.5416
Voltage deviation								
Techniques	TLBO[18]	QOTLBO[18]	MOPSO-1[22]	DE-1[23]	DE	CRO	HCRO	
Best Voltage	0.0913	0.0856	0.2424	0.0911	0.1029	0.0849	0.0841	
Mean Voltage	0.0934	0.0872	NA	NA	0.1035	0.0858	0.0852	
Worst Voltage	0.0988	0.0907	NA	NA	0.1041	0.0867	0.0862	
Voltage stability index								
Techniques	TLBO[18]	QOTLBO[18]	SPEA[19]	DE-1[23]	DE	CRO	HCRO	
Best L index	0.1252	0.1242	0.1397	0.1246	0.1302	0.1156	0.1148	
Mean L index	0.1254	0.1245	NA	NA	0.1310	0.1164	0.1163	
Worst L index	0.1258	0.1247	NA	NA	0.1318	0.1172	0.1162	

**Table 3**  
Comparison of simulation results obtained by different algorithms with STATCOM

Real power loss minimization	Voltage deviation minimization	Voltage stability index minimization
------------------------------	--------------------------------	--------------------------------------

Algorithms	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]
V1 (p.u.)	1.0920	1.0999	1.1000	1.0335	1.0291	1.0293	1.1000	1.0999	1.1000
V2 (p.u.)	1.0822	1.0942	1.0943	1.0316	1.0281	1.0272	1.0994	1.0942	1.0943
V5 (p.u.)	1.0636	1.0747	1.0748	1.0228	1.0232	1.0220	1.0972	1.0747	1.0748
V8 (p.u.)	1.0668	1.0766	1.0767	0.9996	0.9900	0.9997	1.0985	1.0766	1.0767
V11 (p.u.)	1.0998	1.1000	1.0997	1.0160	1.0494	0.9983	1.0994	1.1000	1.0997
V13 (p.u.)	1.0993	1.0953	1.0951	0.9888	0.9973	1.0177	0.9592	1.0953	1.0951
TC6-9	1.0003	1.0387	1.0393	0.9991	1.0063	0.9773	0.9047	1.0387	1.0393
TC6-10	0.9227	0.9000	0.9000	0.9250	0.9405	0.9341	0.9006	0.9000	0.9000
TC4-12	0.9720	0.9724	0.9725	0.9466	0.9696	0.9808	1.0988	0.9724	0.9725
TC27-28	0.9553	0.9628	0.9629	0.9451	0.9446	0.9564	0.9021	0.9628	0.9629
QC10 (p.u.)	0.0469	0.0500	0.0500	0.0300	0.0412	0.0284	0.0483	0.0500	0.0500
QC12 (p.u.)	0.0482	0.0500	0.0500	0.0185	0.0416	0.0000	0.0012	0.0500	0.0500
QC15 (p.u.)	0.0490	0.0500	0.0500	0.0487	0.0484	0.0425	0.0009	0.0500	0.0500
QC17 (p.u.)	0.0489	0.0500	0.0500	0.0008	0.0012	0.0109	0.0487	0.0500	0.0500
QC20 (p.u.)	0.0405	0.0416	0.0419	0.0497	0.0499	0.0454	0.0493	0.0416	0.0419
QC21 (p.u.)	0.0495	0.0500	0.0500	0.0495	0.0350	0.0459	0.0481	0.0500	0.0500
QC23 (p.u.)	0.0262	0.0258	0.0261	0.0490	0.0488	0.0437	0.0499	0.0258	0.0261
QC24 (p.u.)	0.0496	0.0499	0.0500	0.0488	0.0493	0.0457	0.0483	0.0499	0.0500
QC29 (p.u.)	0.0230	0.0225	0.0223	0.0022	0.0063	0.0209	0.0491	0.0225	0.0223
Optimal location	23	23	30	24	23	22	14	23	30
V <sub>STATCOM</sub> (p.u.)	1.0347	1.0454	1.0790	1.0643	1.0553	1.0408	0.9863	1.0454	1.0790
θ <sub>STATCOM</sub> (deg.)	-8.3613	-9.0631	-10.7025	-11.4326	-7.6357	-6.0648	-7.1782	-9.0631	-10.7025
SVD (p.u.)	1.7231	2.0648	2.0869	0.0774	0.0928	0.0803	1.9183	2.0648	2.0869
TL (MW)	4.4812	4.5493	4.5297	5.5232	5.8911	6.1345	7.6840	4.5493	4.5297
L- index	0.1221	0.1311	0.1278	0.1306	0.1403	0.1376	0.1135	0.1286	0.1142

**Table 4**  
Statistical comparison (50 trials) among various algorithms for IEEE 30-bus with STATCOM

Techniques	Real power loss (MW)			Voltage deviation (p.u.)			Voltage stability index		
	HCRO	DE	CRO	DE	CRO	HCRO	DE	CRO	HCRO
Best	4.4812	4.5493	4.5297	0.0928	0.0803	0.0774	0.1286	0.1142	0.1135
Mean	4.4824	4.6106	4.5304	0.0939	0.0807	0.0781	0.1292	0.1154	0.1143
Worst	4.4836	4.7061	4.5332	0.0944	0.0811	0.0788	0.1298	0.1167	0.1151

**Table 5**  
Comparison of simulation results obtained by different algorithms without STATCOM

Algorithms	Real power loss minimization			Voltage deviation			Voltage stability index		
	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]
V1 (p.u.)	1.0597	1.0475	1.0600	1.0179	1.0182	1.0183	1.0157	1.0085	1.0421
V2 (p.u.)	1.0459	1.0333	1.0485	0.9947	0.9927	1.0032	0.9886	0.9789	1.0189
V3 (p.u.)	1.0364	1.0152	1.0365	1.0046	0.9968	1.0034	0.9886	0.9696	1.0138
V6 (p.u.)	1.0267	1.0043	1.0300	1.0017	0.9985	1.0009	0.9869	0.9617	1.0152
V8 (p.u.)	1.0471	1.0279	1.0504	1.0264	1.0229	1.0232	1.0141	0.9863	1.0443
V9 (p.u.)	1.0348	1.0092	1.0321	1.0398	1.0140	1.0167	1.0246	0.9747	1.0089
V12 (p.u.)	1.0331	1.0094	1.0295	1.0063	1.0058	1.0056	1.0108	1.0520	1.0173
TC4-18	0.9008	1.0185	0.9870	0.9922	1.0112	0.9610	0.9573	0.9061	0.9369
TC4-18	0.9038	0.9003	0.9560	0.9891	0.9737	1.0174	0.9884	0.9590	1.0738
TC21-20	0.9880	1.0040	1.0097	0.9765	0.9767	0.9743	0.9808	0.9977	0.9669

TC24-26	0.9916	1.0024	1.0099	1.0640	1.0401	1.0476	1.0701	1.0745	1.0893
TC7-29	0.9005	0.9428	0.9646	0.9497	0.9579	0.9540	0.9343	0.9038	0.9513
TC34-32	0.9740	0.9777	0.9727	0.9096	0.9027	0.9040	0.9095	0.9055	0.9093
TC11-41	0.9001	0.9004	0.9005	0.9003	0.9002	0.9005	0.9006	0.9000	0.9054
TC15-45	0.9005	0.9446	0.9635	0.9772	0.9078	0.9194	0.9143	0.9080	0.9313
TC14-46	0.9001	0.9321	0.9445	0.9406	0.9631	0.9644	0.9810	0.9382	0.9579
TC10-51	0.9119	0.9431	0.9572	1.0126	0.9993	1.0004	1.0012	0.9579	0.9907
TC13-49	0.9003	0.9032	0.9172	0.9009	0.9029	0.9015	0.9012	0.9727	0.9580
TC11-43	0.9011	0.9298	0.9541	0.9645	0.9546	0.9587	0.9602	0.9372	0.9685
TC40-56	1.0149	1.0163	0.9977	1.0394	1.0011	0.9984	0.9965	0.9754	0.9718
TC39-57	0.9834	0.9705	0.9678	0.9014	0.9023	0.9030	0.9220	0.9096	0.9149
TC9-55	0.9067	0.9465	0.9657	1.0158	0.9808	0.9858	1.0058	0.9687	1.0046
QC18 (p.u.)	0.0997	0.0976	0.0953	0.0804	0.0986	0.0737	0.0991	0.0979	0.0924
QC25 (p.u.)	0.0588	0.0588	0.0590	0.0589	0.0590	0.0590	0.0589	0.0580	0.0559
QC53 (p.u.)	0.0630	0.0630	0.0630	0.0629	0.0630	0.0629	0.0629	0.0594	0.0629
SVD (p.u.)	3.7354	1.2007	1.4323	0.6700	0.6919	0.6724	5.1843	5.0365	5.3439
TL (MW)	23.8177	25.1201	24.3835	28.8174	27.8573	27.3553	27.1846	25.1395	24.8609
L- index	0.5523	0.5965	0.5788	0.2898	0.2988	0.2982	0.2236	0.2316	0.2286

**Table 6**

Statistical comparison (50 trials) among various algorithms for IEEE 57-bus without STATCOM

Real power loss												
Techniques	PSO	PSO-w	PSO-cf	CLPSO	SPSO	CGA	AGA	DE	L-SACP-DE	SOA	CRO	HCRO
Best loss (p.u.)	0.2536	0.2597	0.2479	0.2580	0.2742	0.2671	0.2581	0.2512	0.2732	0.2462	0.2438	0.2381
Mean loss (p.u.)	0.2635	0.2839	0.2971	0.2733	0.3070	0.3232	0.2967	0.2618	0.3434	0.2574	0.2443	0.2392
Worst loss (p.u.)	0.2774	0.3249	0.3932	0.3400	0.3862	0.4197	0.3698	0.2730	0.4439	0.2875	0.2451	0.2401
Voltage deviation												
Techniques	OGSA[24]					DE			CRO		HCRO	
Best Voltage deviation (MW)	0.6982					0.6919			0.6724		0.6700	
Mean Voltage deviation	NA					0.6925			0.6731		0.6709	
Worst Voltage deviation	NA					0.6931			0.6738		0.6718	
Voltage stability index												
Techniques	MOPSO-2[25]		MOIPSO[25]		MOCIPSO[25]		DE		CRO		HCRO	
Best L index	0.28834		0.24087		0.23291		0.2316		0.2286		0.2236	
Mean L index	NA		NA		NA		0.2324		0.2293		0.2244	
Worst L index	NA		NA		NA		0.2332		0.2300		0.2252	

**Table 7**

Comparison of simulation results obtained by different algorithms with STATCOM

Algorithms	Real power loss minimization			Voltage deviation minimization			Voltage stability index minimization		
	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]	HCRO	DE[27]	CRO[27]
V1 (p.u.)	1.0407	1.0543	1.0561	1.0277	1.0162	1.0284	1.0045	1.0189	1.0341
V2 (p.u.)	1.0277	1.0415	1.0438	1.0010	1.0016	1.0193	0.9824	0.9941	1.0125
V3 (p.u.)	1.0120	1.0277	1.0362	0.9978	0.9939	1.0157	0.9989	0.9998	0.9995
V6 (p.u.)	1.0019	1.0168	1.0284	0.9938	0.9910	1.0107	1.0011	1.0055	0.9970
V8 (p.u.)	1.0255	1.0368	1.0513	1.0240	1.0224	1.0399	1.0270	1.0329	1.0257
V9 (p.u.)	1.0068	1.0167	1.0322	1.0139	1.0008	1.0134	1.0281	1.0106	1.0411
V12 (p.u.)	1.0059	1.0158	1.0274	1.0019	1.0113	1.0127	1.0300	1.0052	0.9971

TC4-18	0.9019	0.9773	0.9001	1.0154	0.9739	1.0805	0.9844	1.0196	0.9218
TC4-18	0.9008	0.9462	0.9003	0.9306	0.9923	0.9329	0.9970	0.9840	1.0206
TC21-20	0.9948	1.0030	0.9898	0.9780	0.9788	0.9749	0.9766	0.9744	0.9736
TC24-26	0.9869	1.0030	0.9888	1.0320	1.0712	1.0456	1.0594	1.0389	1.0606
TC7-29	0.9011	0.9527	0.9022	0.9567	0.9416	0.9698	0.9515	0.9641	0.9480
TC34-32	0.9731	0.9745	0.9749	0.9022	0.9205	0.9070	0.9121	0.9037	0.9073
TC11-41	0.9001	0.9367	0.9001	0.9007	0.9006	0.9001	0.9006	0.9008	0.9022
TC15-45	0.9025	0.9531	0.9010	0.9477	0.9263	0.9334	0.9312	0.9261	0.9227
TC14-46	0.9018	0.9447	0.9006	0.9108	0.9439	0.9833	0.9885	0.9436	0.9834
TC10-51	0.9162	0.9520	0.9097	0.9953	0.9945	0.9897	1.0190	0.9979	1.0070
TC13-49	0.9011	0.9104	0.9008	0.9071	0.9029	0.9001	0.9000	0.9025	0.9008
TC11-43	0.9005	0.9339	0.9000	0.9748	0.9370	0.9378	0.9360	0.9518	0.9755
TC40-56	1.0162	1.0172	1.0154	0.9818	1.0312	1.0412	1.0959	1.0060	1.0148
TC39-57	0.9808	0.9771	0.9833	0.9212	0.9048	0.9133	0.9001	0.9006	0.9004
TC9-55	0.9029	0.9511	0.9030	0.9797	0.9657	0.9813	1.0020	0.9793	1.0157
QC18 (p.u.)	0.0998	0.0939	0.0985	0.0233	0.0776	0.0399	0.0986	0.0998	0.0239
QC25 (p.u.)	0.0582	0.0590	0.0589	0.0588	0.0587	0.0590	0.0590	0.0588	0.0588
QC53 (p.u.)	0.0623	0.0630	0.0630	0.0627	0.0353	0.0630	0.0630	0.0616	0.0628
Optimal location	42	33	45	16	37	37	33	42	34
V <sub>STATCOM</sub> (p.u.)	1.0613	1.0488	1.0522	1.0378	1.0592	1.0537	1.0454	1.0116	1.0874
θ <sub>STATCOM</sub> (deg.)	-12.1324	-11.2892	-10.1368	-12.0121	-8.3640	-11.1346	-9.0631	-8.3941	-10.0146
SVD (p.u.)	1.7867	1.2012	2.1365	0.6523	0.6803	0.6533	2.0648	1.1559	1.3094
TL (MW)	22.9937	24.2316	23.8378	28.4328	27.1392	26.5963	24.5493	26.1384	24.5802
L- index	0.5721	0.5618	0.5901	0.2941	0.2983	0.2947	0.1186	0.1281	0.1253

**Table 8**

Statistical comparison (50 trials) among various methods for IEEE 57-bus with STATCOM

Techniques	Real power loss minimization			Voltage deviation			Voltage stability index		
	HCRO	DE	CRO	HCRO	DE	CRO	HCRO	DE	CRO
Best	22.9937	24.2316	23.8378	0.6523	0.6803	0.6533	0.1186	0.1281	0.1253
Mean	23.0571	25.6371	23.8904	0.6531	0.6813	0.6540	0.1191	0.1289	0.1259
Worst	23.1032	26.9117	23.9655	0.6543	0.6823	0.6547	0.1197	0.1297	0.1265

## 5. CONCLUSION

In this paper the symphony of STATCOM for optimizing various objectives of optimal installations of IEEE 30-bus and IEEE 57- bus is investigated. It is found that the proposed HCRO strategy require less solution runs to obtain the optimal solution unlike CRO, DE strategy which involves more number of solution runs. Chemical reaction optimization (CRO) prove to be an efficient non-linear optimization technique for solving different types of real world problems of various field of engineering. Improvement of the voltage profile minimization of the transmission loss and voltage stability are considered as the objective function to evaluate system performance and it can be observed that the injecting STATCOM there is a reduction of voltage deviation, transmission loss and voltage stability index of a power system network. CRO is used to find the optimal location of STATCOM for solving optimal reactive power dispatch (ORPD) problem. Also it may be observed from simulation results that computation time required for HCRO in all test cases is lesser than that of conventional DE and CRO method. Thus it is found that HCRO algorithm is robust and suitable for locating and sizing STATCOM devices in power system.

## References

- [1]. Urdaneta AJ, Gomez JF, Sorrentino E, Flores L, Diaz R. "A hybrid genetic algorithm for optimal reactive power planning based, upon successive linear programming," IEEE Trans Power Syst 1999, vol. 14, no. 4, pp. 1292-8,
- [2]. Grudin N. "Reactive power optimization using successive quadratic programming method." IEEE Trans Power Syst 1998, vol. 13, no. 4, pp. 1219-25.
- [3]. Lai LL, Ma JT. "Application of evolutionary programming to reactive, power planning – comparison with nonlinear programming approach". IEEE Trans power Syst 1997, vol. 12, no. 1, pp. 198-206.
- [4]. Kim DH, Lee JH, Hong SH, Kim SR. "A mixed-integer programming. Approach for the linearized reactive power and voltage control-comparison with gradient projection approach". In: International conference on energy management and power delivery, proceed of EMPD, vol. 1, 1998, pp. 67-72.
- [5]. Hingorani NG, Gyugyi L. "Understanding FACTS: concepts and technology of flexible AC transmission systems." New York: IEEE Press; 2001.
- [6]. Galiana FD, Almeida K, Toussaint M, Griffin J. "Assessment and control of the impact of FACTS devices on power system performance". IEEE Trans Power Syst 1996, vol. 11, no. 4, pp. 1931–6.
- [7]. Gyugyi L. "A unified power flow control concept for flexible AC transmission systems". IEE Proc Gener Transm Distrib 1992, vol. 139, no. 4, pp. 323–31.
- [8]. Gyugyi L, Schauder CD, Williams SL, Rietman TR, Torgerson DR. The unified power flow controllers: a new approach to power transmission control. IEEE Trans Power Deliv 1995, vol. 10, no. 2, pp.1085–97.
- [9]. Galiana FD, Almedia K, Toussaint M, Griffin J. "Assessment and control of the impact of FACTS devices on power system performance." IEEE Trans Power Syst 1996, vol.11, no. 4, pp. 1931–6.
- [10]. Kumar A, Chanana S. "New secure bilateral transaction determination and study of pattern under contingencies and UPFC in competitive hybrid electricity markets". Int J Electr Power Energy Syst 2009, vol. 31, no. 1, pp. 23–33.
- [11]. Palma-Behnke R, Vargas LS, Perez JR, Nunez JD, Torres RA. "OPF with SVC and UPFC modeling for longitudinal systems." IEEE Trans Power Syst 2004, vol. 19, no. 4, pp.1742–53.
- [12]. Bhowmik AR, Chakraborty AK, Das PN. "Placement of UPFC for minimizing active power loss and total cost function by PSO algorithm." In: 2013 Int. conf. advanced electronic systems (ICAES). 2013, pp. 217–20.
- [13]. Saravanan M, Mary RSS, Venkatesh P, Abraham JPS. "Application of particle swarm optimization technique for optimal location of FACTS devices considering cost of installation and system load ability". Electr Power Syst Res 2007, vol. 77, no. (3–4), pp. 276–83.
- [14]. Eslami M, Shareef H, Taha MR, Khajehzadeh M. "Adaptive particle swarm optimization for simultaneous design of UPFC damping controllers." Int J Electr Power Energy Syst 2014, vol. 57, pp. 116–28.
- [15]. Shaheen HI, Rashed GI, Cheng SJ. "Application and comparison of computational intelligence techniques for optimal location and parameter setting of UPFC". Eng Appl Art Int 2010, vol. 23, no. 2, pp. 203–16.
- [16]. Gerbex S, Cherkaoui R, Germond J. "Optimal location of multi-type FACTS devices in power system by means of genetic algorithms." IEEE Trans Power Syst 2001; 16(3):537–44.
- [17]. Ghadir R, Reshma SR. "Power flow model/calculation for power systems with multiple FACTS controllers." Electr Power Syst Res 2007; 77:1521-31.
- [18]. Mandal B, Roy PK. "Optimal reactive power dispatch using quasioppositional teaching learning based optimization." Int J Electr Power Energy Syst 2013; 53:123-34.
- [19]. Abido MA. "Multiobjective optimal VAR dispatch using strength Pareto evolutionary algorithm." Vancouver, BC, Canada: IEEE Congress on Evolutionary Computation; 2006.

- [20]. Subburaj P, Sudha N, Rajeswari K, Ramar K, Ganesan L. “Optimum reactive power dispatch using genetic algorithm.” Acad open internet J , 2007, vol.21.
- [21]. Devaraj D, Roselyn JP. “Genetic algorithm based reactive power dispatch for voltage stability improvement”. Int J Electr Power Energy Syst 2010, vol. 32, no.10, pp. 1151-6.
- [22]. Durajra S, Kannan PS, Devaraj D. “Multi-objective VAR dispatch using particle swarm optimization.” Emerg Electr Power Syst 2005; vol. 4, no. 1.
- [23]. Abou El Ela AA, Abido MA, Spea SR. “Differential evolution algorithm for optimal reactive power dispatch”. Electr Power Syst Res 2011, vol. 8, no. 2, pp. 458-64.
- [24]. Shaw B, Mukherjee V, Ghoshal SP. “Solution of reactive power dispatch of power systems by an opposition-based gravitational search algorithm.” Int J Electr Power Energy Syst 2014, vol. 55, pp. 29-40.
- [25]. Chen G, Liu L, Song P, Du Y. “Chaotic improved PSO-based Multi-objective optimization for minimization of power losses and L index in power systems.” Energy Convers Manage 2014; 86:548-60.
- [26]. Dai C, Chen W, Zhu Y, Zhang X. “Reactive power dispatch considering voltage stability with seeker optimization algorithm”. Electric Power Syst Res 2009, vol. 79, pp.1462-71.
- [27]. Susanta Dutta, Provas Kumar Roy, Debashis Nandi. “Optimal location of STATCOM using chemical reaction optimization for reactive power dispatch problem.” Ain Shams Engineering Journal 2016; vol. 7, pp. 233–247.



**Susanta Dutta** did his graduation in Electrical Engineering from Dr. B. C. Roy Engineering College, Durgapur, West Bengal, India in the year of 2004. He obtained his Master of Technology Degree from N.I.T, Durgapur, West Bengal, India with specialization in Industrial Electrical System in the year 2007 and PhD Degree from N.I.T, Durgapur, West Bengal, India. in the year of 2015. In November 2006, he joined as a faculty in the Electrical Engineering Department of Dr. B C Roy Engineering College, Durgapur, West Bengal, India. Presently he is associate professor of Dr. B C Roy Engineering College, Durgapur in the Electrical Engineering Department. and his research area of interest included optimal power flow, ORPD, AGC and CHP.





## A REVIEW ON RECENT DEVELOPMENTS IN THE FIELD OF AMORPHOUS SEMICONDUCTORS

Anil Chamuah<sup>1\*</sup>, Sanjib Bhattacharya<sup>1,2</sup>, Chandan Kumar Ghosh<sup>3</sup>

<sup>1</sup> Dept. of Physics, University of North Bengal, Raja Rammohunpur, West Bengal, Email: [chamuahanil@gmail.com](mailto:chamuahanil@gmail.com)

<sup>2</sup> UGC-HRDC, University of North Bengal, Raja Rammohunpur, West Bengal, Email: [ddirhrdc@nbu.ac.in](mailto:ddirhrdc@nbu.ac.in)

<sup>3</sup> Dept. of ECE, Dr. B.C. Roy Engineering College, Durgapur, West Bengal, Email: [mcet\\_ckg@yahoo.com](mailto:mcet_ckg@yahoo.com)

**Abstract** — In amorphous semiconductors, the atomic order in a long range ceases to exist, but the short range order remains to some extent, giving rise thereby to a band-like structure of electron energy states similar to that of crystalline semiconductors. Amorphous semiconductors in many aspects are prioritized over crystalline semiconductors due to its low cost, easy preparation process and market demand for tailored material properties. Now a days, they have found applications in all areas where previously only crystalline semiconductors are used. In this review paper, a few recent discoveries and developments in the field of amorphous semiconductors are discussed.

**Keywords** — *amorphous semiconductors, chalcogenide glasses, dielectric behavior, electrical properties, temperature dependent, structural properties*

### 1. INTRODUCTION

It is well known to us the current demand of the semiconductor industry which is even going to be increase by many folds in the near future due to the rapid growth and development in the electronics industry. At present almost all of our electronic gadgets, from basic electronic household appliances to missiles and satellites are totally based on semiconductors. The most prominent semiconductor materials until recent years are the single crystal or polycrystalline semiconductor materials such as Si or Ge. But the manufacturing of the base, i.e. the semiconductor wafers, it is a very complex, costly and time consuming process. Here comes the concept of amorphous semiconductors which are a new boon to the semiconductor industry. They are easy to prepare, the production cost is also relatively lower, and for most of the devices it can replace the crystalline semiconductors. Moreover, these amorphous semiconductors properties can be tailored to meet any specific device application.[1]–[3]

### 2. REVIEW WORK

Shiv Kumar Pal et al. in [4] has developed multi-component Se-Te-Sn-Ge (STSG) glassy systems for  $x = 0, 2, 4$  &  $6$  and have performed XRD for structural analysis. AC measurements were carried out in vacuum of  $\sim 10^{-3}$  Torr for a temperature range of 313 K – 338 K and frequency range of 1 kHz – 1 MHz.

Figure 1 shows the XRD patterns for the as-prepared glassy systems where no any distinct peak is seen and in turn confirms its amorphous nature.

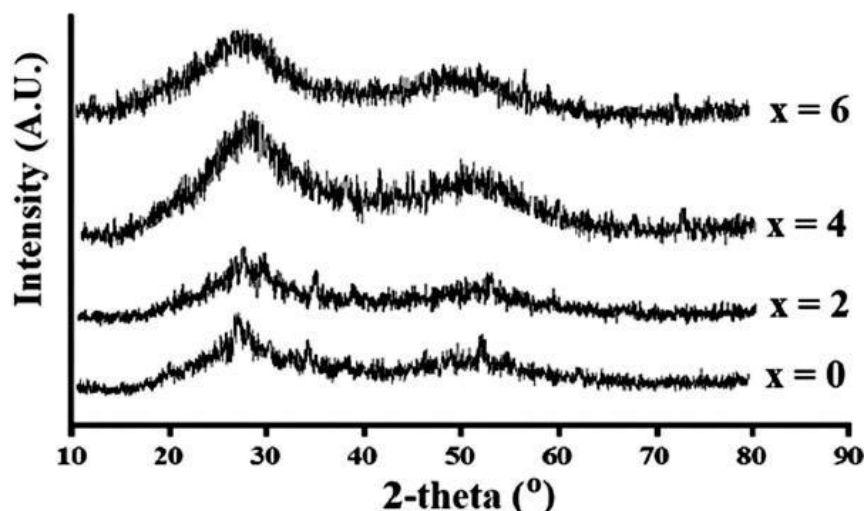


Figure 1 – XRD patterns of Se-Te-Sn-Ge Glassy Alloys for x = 0, 2, 4 & 6

Figure – 2 shows the frequency dependency of the AC conductivity for the as-prepared samples. For x = 0 and 2 the AC conductivity plots are seen to be smooth and increasing with respect to increasing frequency but for x = 4 and 6, random variation is observed which can be understood in terms of the frequency dependency of the dielectric loss for the samples.

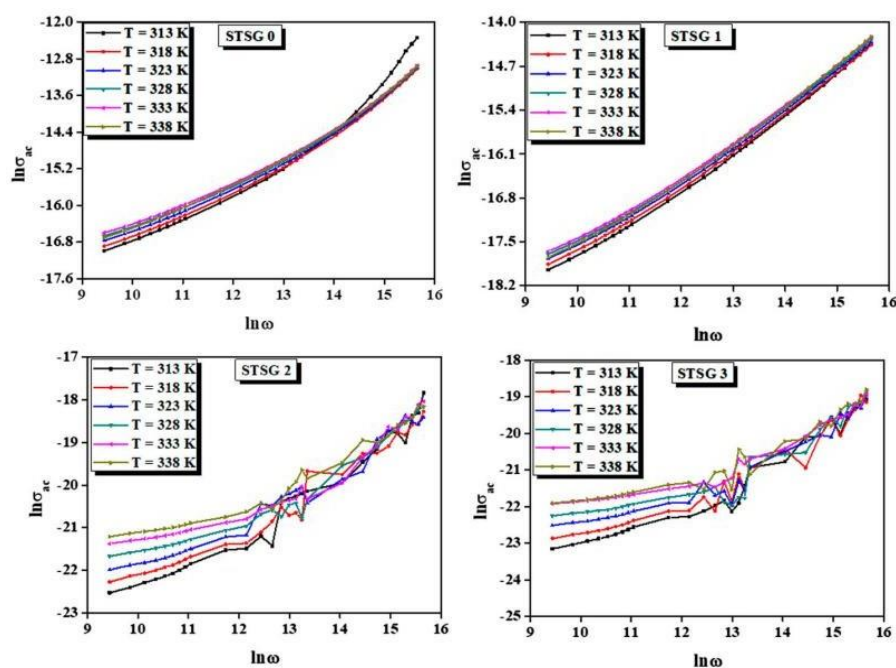


Figure 2 - Frequency dependence of AC Conductivity for the STSG Glassy System

The frequency exponent ‘s’ for all the as-prepared samples are found to be decreasing with increasing temperature and are fitted with CBH model. This glassy system also confirms the applicability of the Arrhenius relation for the thermally activated AC conductivity. The activation energy for the system was found to be decreasing with increasing frequency. As per the Maxwell’s constraint theory, here in this STSG system the observation of drastic

changes in electrical properties such as dielectric loss, AC conductivity, MN activation energy and I-V Characteristics is direct consequence of a stiffness transition. When the value of  $x$  reaches 4, the glassy matrix becomes heavily cross-linked.

Shuo Liu et al. in [5] had fabricated a high-sensitivity, high-detectivity and self-powered UVA photodetector based on polycrystalline MAPbCl<sub>3</sub>/amorphous Ga<sub>2</sub>O<sub>3</sub> heterojunction, as shown in Figure – 3 by the radio frequency magnetron sputtering method.

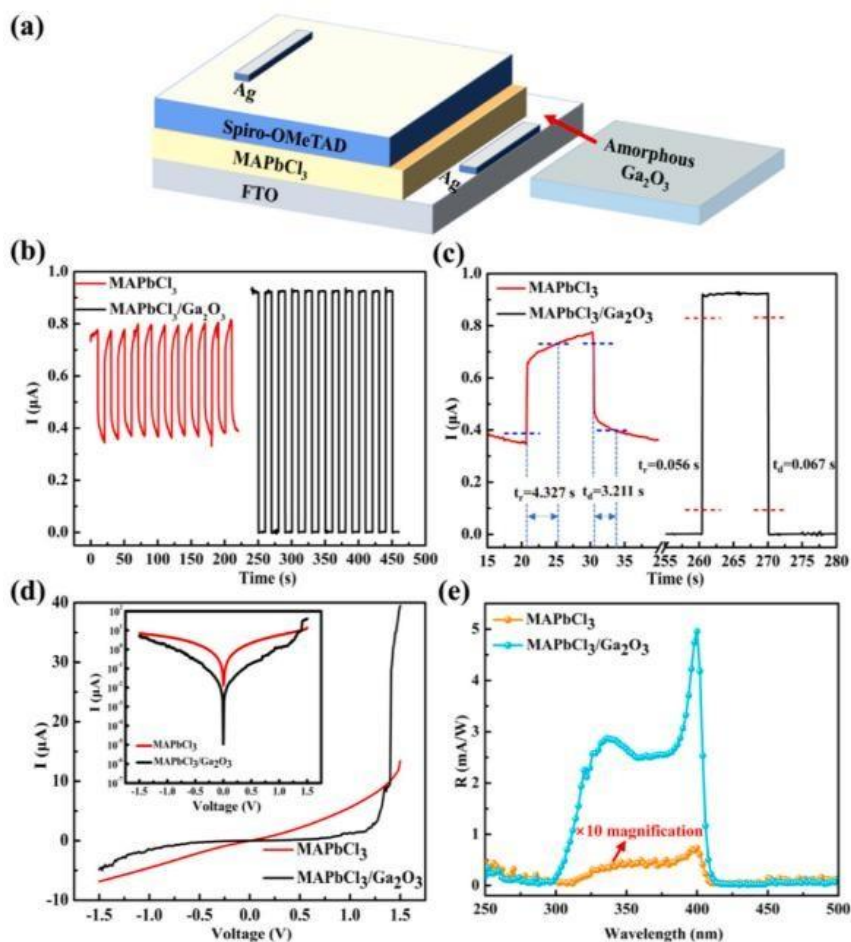


Figure 3 - (a) Schematic diagram of the device structure. (b) I-T curve (c) Rise and Decay Time. (d) Dark I-V curves. (e) Responsivity

The amorphous Ga<sub>2</sub>O<sub>3</sub> layer did not affect the detection range but greatly enhances the performance of the polycrystalline MAPbCl<sub>3</sub> devices. As reported in the literature, the responsivity had increased from  $7.36 \times 10^{-2}$  mA/W to 4.96 mA/W at zero bias. The rise time is shortened from 4.327 s to 56 ms and the decay time has shortened from 3.211 s to 67 ms. The light-dark current ratio had significantly increased from 2 to 890 and the detectivity of the heterojunction detector is as high as  $5.4 \times 10^{10}$  Jones without the application of any external bias supply.

### 3. CONCLUSION

In this review paper, recent discoveries and developments in the field of amorphous semiconductors, their structural and electrical properties have been reported along with their device application and performance parameters.

**Acknowledgement:** The financial assistance from the DST-CRG (Department of Science & Technology, Govt. of India) via Sanction No. CRG/2018/000464 is thankfully acknowledged.

#### References

- [1] P. W. Anderson, "Model for the Electronic Structure of Amorphous Semiconductors," *Phys. Rev. Lett.*, vol. 34, no. 15, p. 953, Apr. 1975.
- [2] D. Adler, "Amorphous semiconductors," *C R C Crit. Rev. Solid State Sci.*, vol. 2, no. 3, pp. 317–465, 1971.
- [3] H. Bark, S. Kim, W. Lee, P. S. Lee, and H. Lee, "Continuous Tuning of the Fermi Level in Disorder-Engineered Amorphous Films of Li-Doped ZnO for Thermoelectric Applications," *ACS Appl. Mater. Interfaces*, p. acsami.1c16162, Nov. 2021.
- [4] S. K. Pal, N. Mehta, and A. Dahshan, "Signature Of stiffness transition in electrical behaviour of Se-Te-Sn-Ge glassy alloys," *Philos. Mag.*, pp. 1–13, Oct. 2021.
- [5] S. Liu *et al.*, "Polycrystalline perovskite CH<sub>3</sub>NH<sub>3</sub>PbCl<sub>3</sub>/amorphous Ga<sub>2</sub>O<sub>3</sub> hybrid structure for high-speed, low-dark current and self-powered UVA photodetector," *J. Alloys Compd.*, vol. 890, p. 161827, Jan. 2022.



**Anil Chamuah** has completed his B. Sc in Electronic Science from Dibrugarh University in 2011, M. Sc in Electronics and Communication Technology from Gauhati University in 2013 and M. Tech in Electronics and Communication Engineering from Rajiv Gandhi University in 2018. He is currently working as a SRF in a DST-CRG Major Research Project in Dept. of Physics, University of North Bengal. His current research is on Chalcogenide Glassy Systems. He had also worked in the field of Microstrip Patch Antennas, MEMS Design, Artificial Neural Networks and Microcontrollers.



**Dr. Sanjib Bhattacharya** is working at University of North Bengal as an Associate Professor-Deputy Director of UGC-HRDC and Adjunct Faculty of Department of Physics. Previously He was an Assistant Professor of Physics at Siliguri Institute of Technology during last 10 years. He is also Principal Investigator of various major research projects, funded by CSIR and DST, Govt. of India. He published 60 papers in reputed international journal, Book and Book chapters and conference proceedings. 3 students have already completed their PhD works under his guidance and many more are working under his supervision. His keen interest is to develop new generation solid state materials for modern battery application, Glass and Glass-nanocomposites and Chalcogenide Glassy Alloys.



**Dr. Chandan Kumar Ghosh** (1964) received the B.Sc (Hons) degree in Physics and B.Tech. degree in Radio physics and Electronics from University of Calcutta in the year 1987 and 1990 respectively. He did M.Tech. degree in Microwave Engineering in the year 2003, from Burdwan University, India. He also received his Ph D in the year 2014 from IEST (Formerly known as BESU), Shibpur, India. From 1991 to 1995, he worked as Development Engineering in Sonodyne Electronics Co. Pvt. Ltd. and from 1996 to 1999 he worked as Assistant Manager (R&D) in Sur Iron & Steel Co. Pvt. Ltd. From 2000 to 2009 he was associated in the department of Electronics & communication Engineering of Murshidabad College of Engineering & Technology and from Aug'2009 he is in Dr. B. C. Roy Engineering College, Durgapur, India. He was in the post of HoD from 2012-2014 in the Dept. of ECE and presently holds the post of Professor. He published more than 60 contributory papers in referred journals and international conference proceedings. His current research interests include wireless communication, Array antenna using microstrip structures, MIMO antenna, DGS, EBG and microstrip resonator's integrated antenna structures.



# OPPORTUNITIES AND CHALLENGES FOR IT INDUSTRIES IN INDIA

Pijush Pal Roy<sup>1</sup>, Chandan Koner<sup>2\*</sup>

<sup>1</sup> Ex-Director, Dr. B. C. Roy Engineering College, Durgapur

<sup>2</sup> Dr. B. C. Roy Engineering College, Durgapur. Email: [chandan.koner@bcrec.ac.in](mailto:chandan.koner@bcrec.ac.in)

**Abstract** — Information technology has given India frightening brand equity in the global markets. Communication infrastructure is also improving in India. The different industrial & trading policies of India are discussed in this paper. The opportunities in different sectors of industry are also given.

**Keywords** — *SEZ, Industrial Policy, Trade Policy*

## 1. INTRODUCTION

Information technology (IT) is defined as, the collection, storage, processing, transmission and presentation of information of any system or sub-systems with the help of electronics, computing and telecommunication devices at the quickest possible time in decision making process or for other purposes. Over the past decade, information technology industry has become one of the fastest growing industries in India. Strong demand over the past few years has placed India amongst the fastest growing IT markets in the Asia- Pacific region. Banking, financial services and insurance, and technology (hi-tech/ telecom) are the main verticals, accounting for nearly 60 percent of the total; manufacturing, retail, media, utilities, healthcare, education, mining, power, transportation and other sectors also growing rapidly.

India offers a unique combination of attributes that have established it as the preferred offshore destination for IT industries. The visibly higher preference for India is driven by its unmatched superiority when measured across a range of parameters that determine the attractiveness of IT industry, those include: abundant trained human resource, cost advantage, emphasis on quality and information security, rapid growth in key business infrastructure etc.

## 2. INVESTMENT POLICY AND REGULATIONS

Information technology has given India formidable brand equity in the global markets. Keeping this in view, the Government has set up a National Manufacturing Competitiveness Council to provide a continuing forum for policy dialogue to energize and sustain the growth of manufacturing industry including IT hardware.

India has been successfully promoting reforms in IT sector. Being a signatory to the information technology agreement of the World Trade Organization, the customs duty on all the specified 217 items has been eliminated, from March 1, 2005.

### A. Industrial Approval Policy

Industrial licensing has been virtually abolished in the electronics and IT sector except for manufacturing electronic aerospace and defence equipment. There is no reservation for public sector enterprises in the electronics and IT industry and private sector investment is welcome in every area. Electronics and IT industry can be set up anywhere in the country, subject to clearance from the authorities responsible for control of

environmental pollution and local zoning and land use regulations.

### **B.Foreign Investment Policy**

A foreign company can start operations in India by registration of its company under the Indian Companies Act 1956. Foreign equity in such Indian companies can be up to 100 per cent. At the time of registration it is necessary to have project details, local partner (if any), structure of the company, its management structure and shareholding pattern. Registration is a kind of formality and it takes about two weeks. It can forge strategic tie up with an Indian partner. A joint venture entails the advantages of established contracts, financial support and distribution-marketing network of the Indian partner. Approval of foreign investments is through either automatic route or Government approval.

\*Corresponding author: Chandan Koner, Email: chandan.koner@brec.ac.in

Foreign technology induction is encouraged both through Foreign Direct Investment (FDI) and through foreign technology collaboration agreement. FDI and foreign technology collaboration agreements can be approved either through the automatic route under powers delegated to the Reserve Bank of India or otherwise by the Government.

### **C.Foreign Trade Policy**

In general, all electronics and IT products are freely importable, with the exception of some defence related items. All electronics and IT products, in general, are freely exportable, with the exception of a small negative list which includes items such as high power microwave tubes, high end super computer and data processing security equipment.

Export Promotion Capital Goods (EPCG) scheme allows import of capital goods on payment of 5 per cent customs duty. The export obligation under EPCG scheme can also be fulfilled by the supply of Information Technology Agreement (ITA-1) items to the Domestic Tariff Area (DTA) provided the realization is in free foreign exchange. The import of second hand computers including personal computers and laptops are restricted for imports.

### **D.SEZ Scheme**

Special Economic Zone (SEZ) is a specifically delineated duty free enclave and shall be deemed to be foreign territory for the purposes of trade operations and duties and tariffs. SEZ unit may import/procure from the DTA without payment of duty all types of goods and services, including capital goods, whether new or second hand, required by it for its activities or in connection therewith, provided they are not prohibited items of imports.

The units shall also be permitted to import goods required for the approved activity, including capital goods, free of cost or on loan from clients. SEZ unit may, on the basis of a firm contract between the parties, source the capital goods from a domestic/foreign leasing company. SEZ unit shall be a positive Net Foreign Exchange earner. Net Foreign Exchange Earning (NFE) shall be calculated cumulatively for a period of five years from the commencement of production.

As per the 'Special Economic Zones Rules, 2006', notified by the Department of Commerce, in case a SEZ is proposed to be set up exclusively for electronics hardware and software, including information technology enabled services, the area shall be ten hectares or more with a minimum built up processing area of 1,00,000 square meters.

### **E.Export Promotion Schemes**

Special schemes are available for setting up Export Oriented Units for the electronics/IT Sector. Various incentives and concessions are available under these schemes. The schemes are:

- Export Oriented Unit (EOU) Scheme
- Electronics Hardware Technology Park (EHTP) Scheme

- Software Technology Park (STP) Scheme
- EOU/EHTP/STP Schemes

Units undertaking to export their entire production of goods and services, except permissible sales in the DTA, may be set up under the EOU, EHTP or STP Scheme for manufacture of goods, including repair, re-making, reconditioning, re-engineering and rendering of services. Trading units, however, are not covered under these schemes.

100 per cent FDI is permitted through automatic route for the units set up under these schemes. These units may import and/or procure from the DTA or bonded warehouses in DTA, without payment of duty, all types of goods, including capital goods, required for its activities, provided they are not prohibited items of import. The units shall also be permitted to import goods including capital goods required for the approved activity, free of cost or on loan/lease from clients.

A unit under any of these schemes may, on the basis of a firm contract between the parties, source the capital goods from a domestic/foreign leasing company without payment of customs/excise duty. This unit shall be a positive net foreign exchange earner. Net Foreign Exchange Earnings (NFE) shall be calculated cumulatively in blocks of five years, starting from the commencement of production.

Supplies of Information Technology Agreement (ITA-1) items and notified zero duty telecom/electronic items affected from EOU/EHTP/STP units to DTA will be counted for the purpose of fulfilment of positive NFE.

The Software Technology Parks of India (STPI) scheme has played a pivotal role in catalyzing the growth of this sector and supporting its rapid proliferation across the country. The tax holiday has helped attract much needed investments (MNC and Indian) in the sector and the virtual model has allowed firms to avail benefits without constraints on their choice of location – encouraging entrepreneurship and integrated growth.

Although the existing term of the STPI scheme is nearing its end (in 2009) the Government intends to continue the benefits offered, by introducing similar provisions in the SEZ policy – and further relaxing the minimum area requirements (to qualify for an SEZ status), for the IT sector..

### 3. OPPORTUNITIES AND CHALLENGES FOR IT SECTOR

#### A. Education

More than three decades ago, computers and related information technologies were introduced to educators as educational tools. The interactive IT solution provides both traditional and distance learning institutes simple and logical assistance when developing and scheduling courses, planning activities and most significantly, improving collaboration and communication between teachers and students. Traditional lecture methods are often left behind as students collaborate and teachers facilitate with IT.

#### B. Opportunities

Application of IT has following opportunities in education sector in India:

*Advanced skills development:* The new technologies are required for students to have more control over their own learning, to think analytically and critically, and to work collaboratively. This will help them to do research, engage in projects, and to communicate. This will make a reform in educational system.

*Administrator tools:* Administrators requires computer and information technologies to improve their roles in the educational process. They can use internal report generator and statistics tool for student

registration and enrolment, staff registration, scheduling purpose and inventory registration.

*Teachers' tool:* IT application is required for teachers for generating built-in communication tools, providing a collaborative environment in which student and teachers can easily and effectively work together. Using this tool, teachers can able to take part in portfolio assessment methods and leverage automatically generated reports to quickly and efficiently monitor student progress and follow-up on this individual progress; thereby helping to free up time for instruction or professional development.

*Technology adaptation:* Development of on-line learning module is essential for actively engaged in a constructive process of knowledge sharing. The Problem-based Learning (PBL) approach can be used in education by centring on a problem rather than within the boundary of a subject discipline, learners can acquire necessary knowledge and skills through exploration, problem solving and team work. Students can develop ability to teach others, and can have greater problem solving and critical thinking skills. A user -friendly tool, Adobe Captivate 2, can be used to develop online interactive teaching and learning materials with ease and which help to promote the learning interest and self-learning skills of students. Not only Computer Based Training (CBT) and Web Based Training (WBT) but Graphic and Video Tools, Wireless and Mobile Technology, SMART Boards and ACTIV Boards are other tools of IT can be used more in education. Further Smart boards can be integrated with wireless PDA.

### **C. Challenges**

Success of usage of IT in education sector requires a detailed technology plan and such plan should consider funding, installation and integration of equipment, ongoing management of the technology. The plan should also express a clear vision of the goals of the technology integration. Teacher training is the main requirement. Teachers should know how to operate the technology and how to integrate it into the curriculum. Administrative support is required in the form of funding, or in restructuring schedules and physical space to reflect the new learning environment. Parents, businesses, and community members can use technology as a springboard to become more involved in the activities of institutes. All can help with wiring or technical support. Parents can use e-mail to facilitate communication with teachers and administrators. Businesses can use e-mail to help mentor students and help them prepare for the workplace. Support from government is required so that it provides adequate funding and appropriate policy making to assure that technology is accessible to all institutes on an equal basis. If these challenges are met then technology becomes not only an educational tool but an integral part of entire instructional process.

### **D. Health Care**

IT in health is processing the information and applying it to the health care. The idea is actually the re-design of the process of health care to take advantage of the technology. Many current health crises like obesity, hypertension, malaria and respiratory illness are preventable if dealt with early. Early intervention is a far better economic strategy, and may be the only route to sustainable health in future.

Improved automation of medical records, new sensing technologies, wide availability of mobile technologies, and improved treatments are all helping mitigate the health challenges to come. New technologies make possible a variety of health improvements including environmental sensing, remote diagnosis, and health education. Often these technologies are quite simple and the emphasis is on cheap, sustainable solutions.

Researchers are creating new ways to improve and deliver healthcare using IT, from developing tiny monitors that can remotely detect pathogens, diagnose diseases and monitor the state of a patient to dealing with the issues of data privacy and security of electronic medical records.

### **E. Opportunities**

In addition to containing costs, advanced information technologies furnish health care providers with the opportunity to improve patient care by streamlining clinical processes and creating a seamless flow of information. Currently, health care providers use paper-based records to record a patient's receipt of health care services. Unfortunately, the use of such records leads to the inadequate documentation of the care-giving process, a severe disruption in the flow of patient related information, and a substantial delay in the delivery



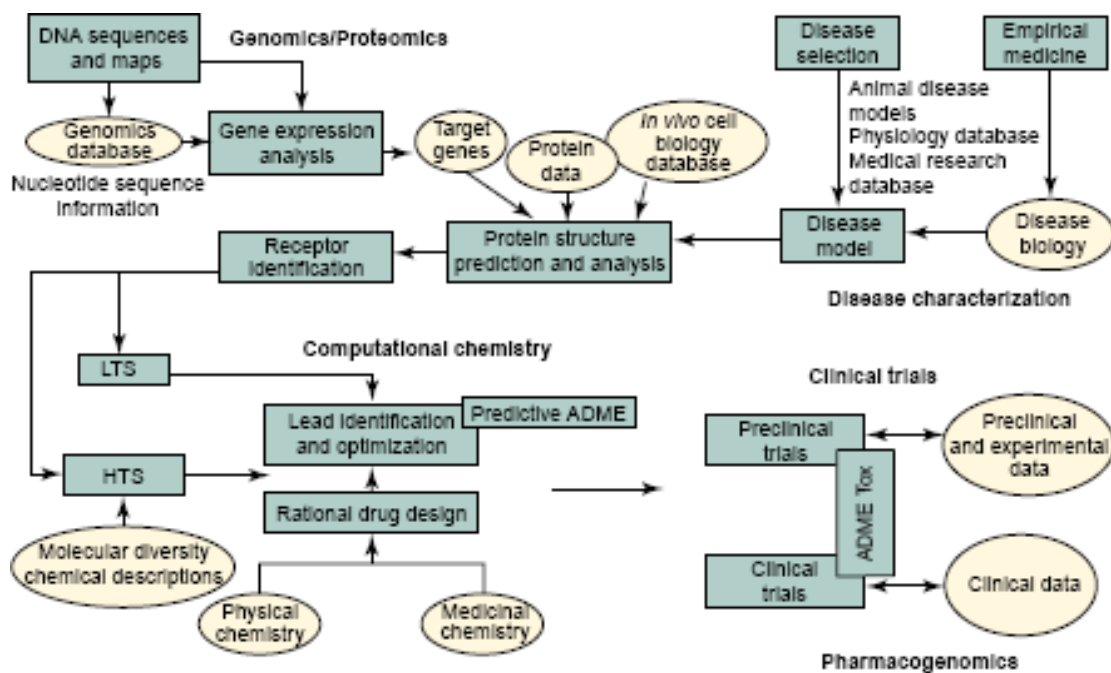
of health care services. Advanced information technologies - such as computer-based patient records, portable computers, and expert information systems - alter this situation by providing clinicians with real-time access to patient information at the point of care. The opportunities for IT application in health care are as follows:

*Health care at home:* "Smart home" technologies can help improve health in many ways from remote check-ups to long-term monitoring to recognizing crises. Implementation of sensor networks that will automatically alert emergency workers via a cell phone when a person falls. With proper security and privacy precautions, the data eventually gathered by these types of sensor networks, such as heart rate and blood sugar, could be made available to doctors to use for everything from diagnostics to research.

*Mobile and worn devices:* Some very inexpensive devices are appearing that allow continuous monitoring of chronic health problems. Mobile devices can also function as "wellness advisors" as people make everyday choices that affect their health. Medical PCs, PDAs, and Portable Workstations are part of IT infrastructure of any modern hospital. As hospitals build up their IT infrastructure, some are installing bedside computers as an alternative to mobile workstations. That way, clinicians do not have to go far to look up charts, enter orders, and document encounters, and while patients can use the same computers to surf the Internet or e-mail friends and family

*Telemedicine:* Healthcare is advancing so rapidly that it is extremely difficult for doctors to manage huge patients. Therefore, a IT solution is required for: (i) Early detection of infectious disease outbreaks around the country; (ii) Improved tracking of chronic disease management; (iii) Consumer empowerment to create a consumer-directed and secure electronic health-care registration information and medication history for patients; (iv) Chronic care to use secure messaging, such as email, for communication between patients and their health-care providers; (v) Electronic health records to create standardized, secure records of past and current laboratory test results that is accessible by health professionals; and (iv) Bio-surveillance to enable the transfer of standardized and anonymized health data to authorized public health agencies within 24 hours.

*Drug discovery:* During the past 25 years, the drug discovery process and a variety of information technologies have co-evolved to the point where they have become inseparable components of a pipeline that begins with basic research and ends with disease specific pharmaceuticals. The impact of this trend has been tremendous with regard to acceleration of the drug discovery process. Information technologies for chemical structure prediction, heterogeneous database access, pattern discovery, and systems and molecular modelling have evolved to become core components of the modern drug discovery process. Rapid advances in the industrialization of gene sequencing combined with databases of protein sequence and structure have created a target-rich but lead-poor environment. During the next decade, newer information technologies that facilitate the molecular modelling of drug-target interactions are likely to shift this balance towards molecular-based personalized medicine – the ultimate goal of the drug discovery process (Figure-1).



**Fig. 1:** Flow of information through the drug discovery pipeline (Augen, 2002)

## F. Challenges

There should be well made policies to address privacy and security issues for nation-wide use of health IT. Clinical decision support is required to improve role of health-care quality. A plan is required for standardized integration of test information into electronic health records. While a larger number of health care practitioners are becoming fascinated with IT and systems analysis, many are still reluctant to accept it. They are not trained on the computer to gather information and make decisions, so it's difficult for them to change the way they practice. So, well planned training program is the main challenge.

### 3.1. Power Sector

IT is widely acknowledged to be crucial for efficient operation and management of all industrial systems. This is equally true of the power utilities, which need to handle a large amount of information for their efficient operation. IT is one of the most promising industrial sectors of the Indian economy and India's IT capability is well recognized globally. At the same time the opening and restructuring of the Indian power sector have changed the perception of the power utility managers in the way they have been doing business till now. The entire gamut of the operation in power sector will be revolving around better availability of quality power and efficient management of energy while maintaining the eco-friendliness all along.

#### 3.1.1. Opportunities

Information technology has many scopes for application in a wide spectrum of tools ([www.powermin.nic.in](http://www.powermin.nic.in); [www.teil-india.com](http://www.teil-india.com); [www.walawalkar.com](http://www.walawalkar.com)) that will enhance performance in planning, operational, maintenance, training and commercial activities in a power system (Figure-2). These include-

- Monitoring various online real-time plant parameters from different generating stations;
  - Establishing a full-fledged generation control room for performance diagnosis and optimization; and
- Providing suitable connectivity between each power station and control room.

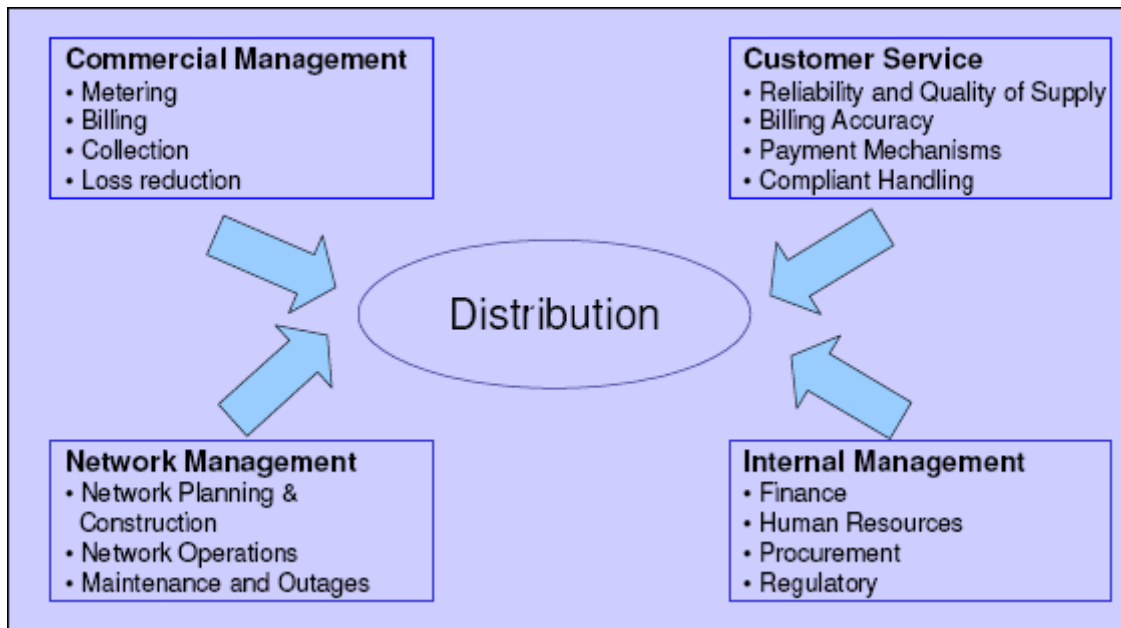


Fig. 2: Functional areas and processes in power sector (Gupta, 2003)

Specific information systems in power sector include:

- Energy management system
- Distribution automation system
- Planning and energy accounting system
- Metering and billing system
- Training system
- Maintenance management system
- Management information system
- Power plant control system
- Supervisory control and data acquisition system
- Generation, energy efficiency and monitoring system
- Transmission, including smart grid, open access, power trading, energy exchange, substation automation and integration
- Distribution covering performance improvement, asset and system management, geographical information system, smart metering, revenue protection, energy audit, customer indexing, trouble call management.

### 3.1.2. Challenges

The power sector is opening up and use of IT and automation systems can play a major role in bringing about standardization, transparency, revenue realization and reduction in transmission and distribution losses in the sector. Technology innovation can only benefit the sector and IT has a major role to play in empowering the power supply utilities.

There is a huge need for specialized, customized and upgraded software systems for the power sector.

The size and the scope of the software order are significant. India's state-owned power utilities are expected to place significant orders of software that can help them in project monitoring and in managing their networks better.

### 3.2. Coal Mining

Indian coal mining industry is continued to face new and greater challenges in the post-liberalization era to reduce production cost, improve productivity, increase safety, improve environmental conditions, increase mechanization and enhance efficiency to meet the demand of coal. In the present situation, it has become imperative to adopt IT, which facilitates better management and cost-effective performance of highly capital-intensive projects for enabling coal mining industry to remain viable and competitive.

#### 3.2.1. Opportunities

Application of IT in mining industry has huge potential in the following areas (Chaulya et al., 2008):

**Improving production and productivity:** Real time measurement of production using sensor and computer, i.e. on-line data feeding from the dispatch conveyor or the tippler where the raw coal is dumped, can make the mine manager pro-active instead of doing the production analysis and take decision at the shift end or the day after.

**Shift and personnel management:** It will give new impetus to major functions, like (a) on-line monitoring of data, (b) day-to-day planning, (c) resource management and (d) face quality control.

**Reduction of production discrepancy:** An on-line measurement using sensor will be implemented at all the mines. The miners will be paid through the weight-o-meter reading, the discrepancy will be reduced and the mine will have to pay the revenue as per their actual production in the weight-o-meter.

**Maintenance of equipment:** A computerized maintenance management of equipment module is required to overcome the existing problems.

**Management of inventory:** The computerized system will reduce the store cost of each mine and the mining company as a whole will be benefited. If the material list of all the stores of a company are in a web then any mine can draw the material from any store of other mine (essentially from the same company) so that the non-moving item of one mine can be utilized by other.

**Wireless environmental monitoring:** To overcome the existing problems, on-line sensor based environmental monitoring system will be integrated with the proposed system for continuous monitoring of vital environmental parameters in normal condition. The system provides on-line visual representation of trend of all the monitored parameters and gives audio-visual warning signal when a particular parameter crosses the respective threshold limit so that mine management can immediately take appropriate action.

**Disaster forecasting:** The system will provide early audio-visual warning signal to forecast different category of disasters using special sensors and techniques so that mine management can immediately take appropriate action.

**Mine safety management:** Using the computer data base, all the safety and rescue related data can be uploaded in the system for future use. Application of IT can help the mine management through storage of detailed information of accidents and extraction of meaningful data or analysis of data to identify the possible areas of weakness in the mines safety system, which can be used as a guideline for the decision making process to improve mine safety performance

**Statutory requirements:** All the statutory related data, acts, rules, regulations and standards could be fed into computer for accessing the required information as and when required.

**Post-disaster management:** The computerized system will be developed to provide accident analysis, the accident prone areas, probable remedies, emergency response plan, list of first aid, rescue trained personnel along with their address, telephone numbers and place of duty during an emergency. Mobilization of manpower and resources can be done effectively to rescue the trapped miners without any delay.

**Improvement in working by on-line record keeping:** A computerized database for different purposes will be prepared for regularly feeding of data.

**Wireless communication in underground mine:** With the advent of IT, it is possible to implement a proper and reliable wireless communication system in underground mines, which not only save machine break down time, but also help in immediate passing of message from the vicinity of underground working area to the surface for speedy rescue operation

**Decision making:** An appropriate audio-visual communication links will be established for effective, fast and instant decision-making at different management levels and keeping history of important decision in database for future reference.

**Training:** Various training modules will be made available in the proposed IT system for implementation in the training programmes. Identification of the need for training will be easy through computerized database. With computerized the training centre, the technical trainers will be relieved of doing unproductive work.

### 3.2.2. Challenges

The major challenges for IT industry are to develop an integrated system by incorporating application software, networking software and hardware system having following facilities:

- Different modules for data entry, logging, on-line monitoring, processing, analysis, storing and transmitting.
- Web-page for dissemination of mining information, information exchange, video-conferencing and decision-making utilities for effective, fast and instant decision-making at different management levels.
- Decision support system with mining data base.
- Network architecture for connecting mining sites with area and head offices and providing communication strategies.

### 3.3. Banking Services

Information technologies and innovations are the strategic tools for enhancing the value of customer relationship. They reduce the costs of financial transactions, improve the allocation of financial resources, and increase the competitiveness and efficiency of financial institutions. Technology has opened up new markets, new products, new services and efficient delivery channels for the banking industry. Online electronics banking, mobile banking and internet banking are just a few examples. Information Technology has also provided banking industry with the wherewithal to deal with the challenges the new economy poses. Information technology has been the cornerstone of recent financial sector reforms aimed at increasing the speed and reliability of financial operations and of initiatives to strengthen the banking sector.

The IT revolution has set the stage for unprecedented increase in financial activity across the globe. The progress of technology and the development of world wide networks have significantly reduced the cost of global funds transfer. It is information technology which enables banks in meeting such high expectations of the customers who are more demanding and are also more techno-savvy compared to their counterparts of the yester years. They demand instant, anytime and anywhere banking facilities. IT has been providing solutions to banks to take care of their accounting and back office requirements. This has, however, now given way to large scale usage in services aimed at the customer of the banks. IT also facilitates the introduction of new delivery channels - in the form of Automated Teller Machines, Net Banking, Mobile Banking and the like. Further, IT deployment has assumed such high levels that it is no longer possible for banks to manage their IT implementations on a stand alone basis with IT revolution, banks are increasingly interconnecting their computer systems not only across branches in a city but also to other geographic locations with high-speed network infrastructure, and setting up local area and wide area networks and connecting them to the Internet. As a result, information systems and networks are now exposed to a growing number. Technology products in banking sector are (Subbarao, 2009): (i) Net Banking, (ii) Credit Card Online, (iii) One View, (iv) InstaAlerts, (v) Mobile Banking, (vi). NetSafe, (vii) e-Monies Electronic Fund Transfer, (viii) Online Payment

of Excise & Service Tax, (ix) Phone Banking, (x) Bill Payment, (xi) Shopping, (xii) Ticket Booking, (xiii) Railway Ticket Booking through SMS, (xiv) Prepaid Mobile Recharge, (xv) Smart Money Order, (xvi) Card to Card Funds Transfer, (xvii). Funds Transfer (eCheques), (xviii) Anywhere Banking, (xix) Internet Banking, (xx) Mobile Banking, (xxi) Bank@Home, (xxii) Cash on Tap: (a) Express Delivery and (b) Normal Delivery etc.

### **3.3.1. Opportunities**

The potential of IT for the near future includes:

- Extending core banking, mobile banking and ATM facilities in rural areas using solar based power system and VSAT.
- Enabling differentiation in customer service;
- Facilitating Customer Relationship Management (CRM) based on available information, which can be stored and retrieved from data warehouses;
- Improving asset-liability management for banks, which has a direct bearing on the profits of banks;
- Enhancing compliance with anti-money laundering regulations; and
- Complying with Basel II norms.

### **3.3.2. Challenges**

New technologies set off a process of change. That, in turn, poses its own set of challenges to institutions as well as to consumers. IT is not yet a very comfortable choice for millions. Therefore, if we are to encourage IT proliferation, we must facilitate a change in customer mindsets and attitudes. Consumer awareness is a major challenge. It must be addressed as a whole. Security in an IT-based transaction processing environment is also critical. Adequate security controls must be in place. This includes the validation of transactions using the maker-checker concept, transmission of electronic messages over a network in encrypted form, due authentication by means of providing for digital signatures and storage of electronic records in conformity with the provisions of the IT Act, 2000 and amendment Act 2008.

## **3.4. Electoral Process**

Conduct of elections in India is an event that involves mammoth complexity and intensive planning considering electorate size, geographical spread and terrain of India. The advent of IT in Indian Elections has once again underlined the fact, how technology is a boon to the entire mankind (Cardoso, 2011). The Election Commission of India is on a mission to integrate ITs in the Indian electoral process within constitutional provisions and a commitment to hold regular, free and fair elections.

### **3.4.1. Opportunities**

The IT activities in electoral process may be classified as:

#### **a) Pre-election activities**

- Electoral rolls management system
- Preparation of electoral rolls and EPIC
- Drafting of polling personnel/ party, randomization of EVMs and micro observers
- Know your polling booth application - SMS based, Web based, IVRS based, Voice based, and Voter Slip
- Implementation of GENESYS

#### **b) Election-day activities**

- Communication plan for Election Tracking (ComET)
- SMS based services to monitor election process milestones and polling progress
- Web casting/ video streaming of poll proceedings from polling stations

**c) Post election - results/trends**

**d) Other activities related to elections**

- Design and hosting of CEO website
- Hosting of electoral rolls
- Website for redressal of public grievances
- Customized web portal
- Network support for data and video requirements
- Setting up of media centre for print and electronic media at Counting location
- GIS for election management, rationalization of polling stations,
- Election planning and display of election results.

### **3.4.2. Challenges**

The challenges for IT solutions provider for electoral process are as follows:

- Providing enhanced and standardized tools for preparation and maintenance of electoral roll,
- Optimizing operations by effective use of ICT,
- Providing timely information for easier and faster analysis,
- Increasing operational efficiency by removing non-standardized and redundant data,
- Automating back-office functions by minimizing human intervention,
- Ensuring transparency and traceability in electoral roll process and data,
- Maintaining high integrity of electoral roll data and process, and
- Improving service delivery to the stakeholders

### **3.5. Indian Railway**

Indian Railway has been a pioneer in the use of IT in India. IT was first adopted in the 1960s, when computerized passenger and freight revenue accounting, operating statistics, payroll, and inventory management were introduced using IBM 1401 computers placed in Zonal data centers. Later on, a notable success has been the introduction of the Passenger Reservation System (PRS) in the mid-eighties, and its subsequent enhancement to a fully networked application. IT is the only enabling method available that can allow for the most cost-effective optimization of Railway resources.

#### **3.5.1. Opportunities**

To reach the stage of rapid expansion of IT systems in the Railways, a backbone of IT resources is required to be set up. The backbone comprises of the components (Fig. 3) mentioned below.

**Train operations:** Train control, Control charting, Train monitoring, Tracking of rolling stock, Train timetabling and scheduling, Accident management, Re-routing and re-scheduling, Rout capacity determination etc.

**Commercial functions:** Passenger (Passenger reservation, Ticketing, Passenger movement analysis); Freight (Goods booking, Consignment tracking, Freight traffic analysis; Disputes and claims).

**Rolling stock production:** Design; Managing production (Production planning and control, Materials management, Human resources management); Costing; Sales and product distribution (Warranties, Positive recall of parts, External sales and exports, Sale documents, Invoice generation and payment realization).

**Rolling stock maintenance:** Locomotives; Passenger stock; Freight stock; Special rolling stock; Track machines; OHE maintenance cars; Workshop management.

**Fixed asset maintenance:** Track maintenance (Maintenance schedules and monitoring, Materials and scrap, Track parameters, Failure trends); Bridge maintenance (Maintenance schedules and monitoring, Materials and scrap, Bridge parameters, Failure trends); OHE maintenance (Signalling maintenance; Buildings and other civil assets).

**Managing Railway projects:** Feasibility studies; Survey.

**Designing fixed structures:** Track formation; OHE; Station building, platforms; Yard layouts; Signalling layouts; Project monitoring.

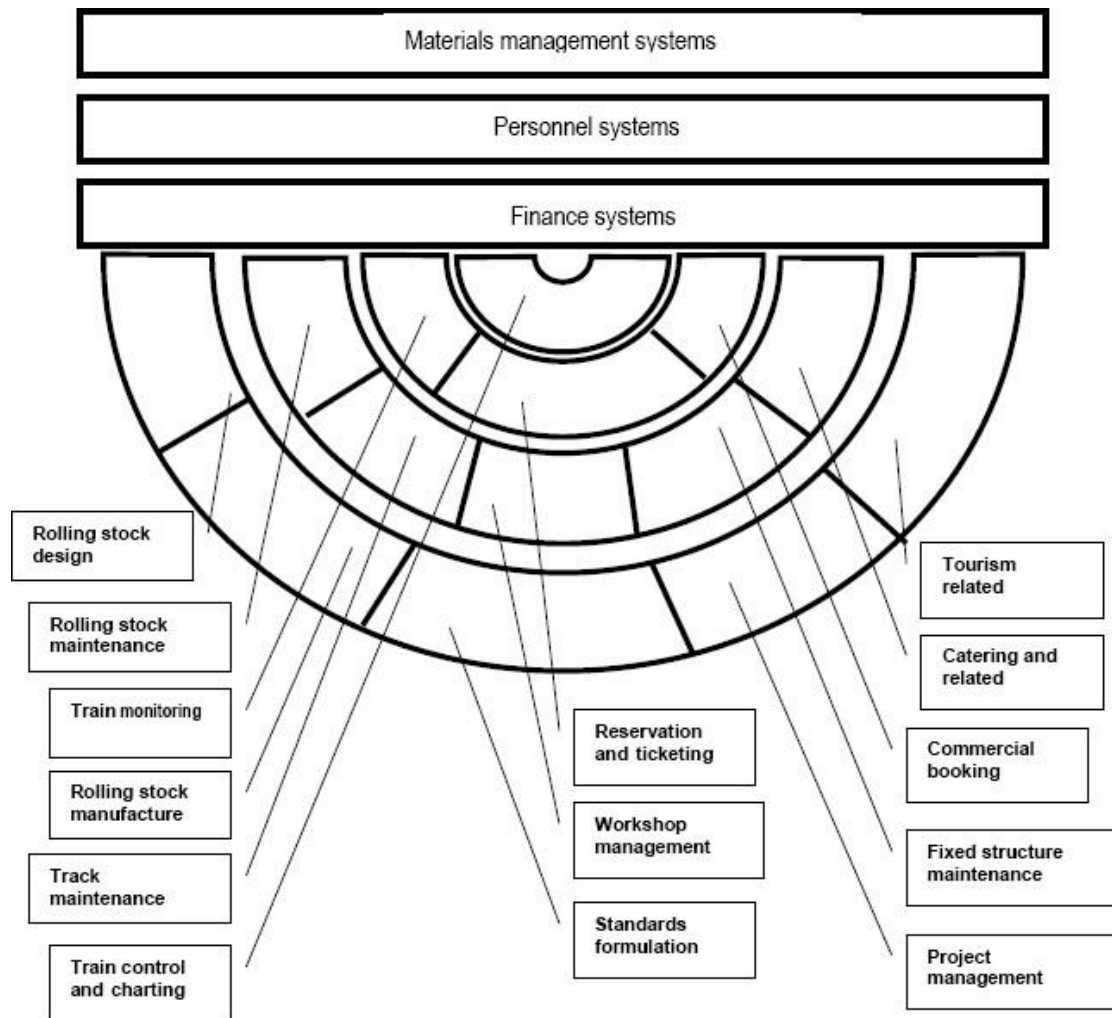
**Personnel management:** Recruitment; Employee payment system (Employee leave accounting, Retirement benefits); Training; Service records; Cadre management.

**Materials management:** Product lifecycle management; Materials database; Procurement and materials sourcing; Scrape management.

**Financial management:** Budgeting and expenditure control; Accounting; Leasing financing; Pre- and Post-financial analysis of projects / investments; Coasting of activities and operations; Assets management.

**Railway Security:** Securing personnel access; Managing criminal databases; Wagon seal monitoring.





**Fig. 3:** Components of information architecture in Indian Railways (Mathur, 2002)

### 3.5.2. Challenges

Security of data and applications are vital challenge as strategic functions begin to depend increasingly on information system. Security starts with physical security of data centers and physical data stores to prevent unauthorized entry and damage to them. Data has also to be secured against program faults and inadvertent or malicious deletion. Remotely located backup should be kept to minimize the impact to damage to the site in the event of a mishap. Similarly, application software may be attacked by inadvertent or malicious modifications, or attacks over the network or through the intranet or internet, depending on the connectivity provided. As systems become more sophisticated, authentication and non-repudiation of messages to the database becomes important. Finally, the system should be auditable, so that any attack on the programs and data can in worst case be detected after the fact.

## 4. CONCLUSION

Over the past decade, information technology industry has become one of the fastest growing industries in India. Strong demand over the past few years has placed India amongst the fastest growing IT markets in the Asia-Pacific region. India offers a unique combination of attributes, namely talent, quality consciousness, cost advantage, vast network of academic infrastructure, young demographic profile, rapid growth etc. which have established it as the preferred offshore destination for IT. The visibly higher preference for India is driven by its unmatched superiority when measured across a range of parameters that determine the attractiveness of IT destination.

**Acknowledgements:** The authors express their sincere thanks and gratitude to Dr. S. K. Chaulya, Sr. Scientist, CSIR-Central Institute of Mining & Fuel Research, Dhanbad for immensely helping during the preparation of the paper. The opinions are of the authors and not the institute to which they belong.

## References

- [1] Augen, J. (2002), The evolving role of information technology in the drug discovery process. *Drug Discovery Today*, 7(5): 315-323.
- [2] Cardoso, F. (2011), Information and communications technology in elections in India. *Informatics*, 20(1): 4-8.
- [3] Chaulya, S.K., Bandyopadhyay, L.K. and Mishra, P.K. (2008) Modernization of Indian coal mining industry: Vision 2025. *Journal of Scientific & Industrial Research*, 67: 28-35.
- [4] Gupta, S. (2003), Application of information technology to improve performance in power. *Power Reforms: Technological and Financial*, Indian Institute of Technology, Kanpur, Sept. 4-5, 2003 ([www.3inetwork.org/workshop/power/Papers/powerreformSanjay.pdf](http://www.3inetwork.org/workshop/power/Papers/powerreformSanjay.pdf)).
- [5] Mathur, S.S. (2002) Vision 2025: technology for information. *Seminar on Indian Railways – Vision 2025*, Organized by Railway Staff College, Vadodara, 5<sup>th</sup> June 2002, pp. 1-10.
- [6] Subbarao, D. (2009), Information technology and banking – a continuing agenda. Reserve Bank of India on May 18, 2009 at the Banking Technology Awards 2008 of the Institute for Development and Research in Banking Technology, Hyderabad, pp. 1-10.



# A Time-Domain Study of the Urban Consumer Price Index of Fuel in India Concerning the Pandemic Period

Paragkanti Chattopadhyay<sup>1</sup>, Uttam Dey<sup>2</sup>, Pranab Sinha<sup>3</sup>, Niloy Kumar Bhattacharjee<sup>4\*</sup>

<sup>1</sup>Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [paragkanti.chattopadhyay@bcrec.ac.in](mailto:paragkanti.chattopadhyay@bcrec.ac.in)

<sup>2</sup>Academy of Professional Courses, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [deyuttam0@gmail.com](mailto:deyuttam0@gmail.com)

<sup>3</sup>Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [pranab.sinha@bcrec.ac.in](mailto:pranab.sinha@bcrec.ac.in)

<sup>4</sup>Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [niloy.bhattacharjee@bcrec.ac.in](mailto:niloy.bhattacharjee@bcrec.ac.in)

**Abstract** — A perennial issue in the economic space is appraised debate about the nature of inflation and magnification of the Indian economy, with particular references being made to the disruptions during the pandemic. India's national income aggregates showed steady magnification in the face of economic downturns in major markets a decennium ago. But the pandemic has taken a heftily ponderous toll on the national economy. An attendant irritant is the truculent inflation trends in critical segments such as fuel. In the present study, we take up a conservative but methodical data science approach to analyze the Urban Consumer Price Index (CPI) of Fuel in India. We study the standard published time series of genuine index values. We run ADF tests with a preliminary ACF/PACF calculation and then tweak the model towards optimization with a formal ARIMA structure in RStudio. In concluding, we make rigorously informed discernments regarding the model's predictive capacity and conclude.

**Keywords** — *ACF, ADF, CPI, ARIMA, Forecast, PACF.*

## 1. INTRODUCTION

The world cumulated to a tricky legitimacy on 11th March 2020 when the World Health Organization (WHO) articulated the Covid19 pandemic[1][2]. Towards the best pandemic disclosure, extraordinary mazuma cognate signs from sundry channels, weeks sooner, showed the world was leaning towards a peculiar watershed for the length of our life, if not in humanity's strategy of encounters[3][4]. In a kineticism of life-transmuting reports, experts across capable structures had since a drawn-out period in the past expected a general pandemic would strain the components of the ecumenical stock chains and conclusively orders, thus genuinely coming to a cross-line monetary fiasco taking into account the incomprehensible interconnect we inhabit. In endeavors to isolate cases and breakpoint the transmission whirlwind of the contamination while mitigating the pandemic, countries across the globe executed crazy evaluations, for instance, virtual public lockdown and inhibit terminations. The monetary shock climbing up out of this pandemic is assessed[5][6].

\* Corresponding author: Niloy Kumar Bhattacharjee, [niloy.bhattacharjee@bcrec.ac.in](mailto:niloy.bhattacharjee@bcrec.ac.in).

A perennial issue in the economic space is informed debate about the nature of inflation and growth of the Indian economy, too, with particular references being made to the disruptions during the pandemic. India's national income aggregates showed a steady increase in the face of economic downturns in major markets a decade ago [7][8]. But the pandemic has taken a heavy toll on the national economy. An attendant irritant is the aggressive inflation trends in critical segments such as fuel.

Monographs, books, and papers on the CPI, the GDP, and other indicators of national income, inflation, employment - trajectories, trends, inflection, cycles, and so on are, literally, a dime a dozen. An incisive compendium on the techniques adopted by various research articles are insights on India's growth turnaround [9]. They appear to lend credence to the curious observation that Indian growth has "arrived" by defying worldwide trends, so economists are merely left the task of fitting explanatory equations to it. They also assert that the GDP growth spurt preceded the adoption of LPG practices in India. Hence, it can hardly be called the result of globalization to exclude other factors. Given the economy's reasonably insulated status during the sub-prime mortgage crisis a decade ago, these assertions would seem to hold water in hindsight, too. The experiment and prediction specific to inflation CPI modeling are also relatively abundant; as an example, ARIMA and ARDL treatment of price movements, inflation trends, and growth [10][11][12].

Sundry modalities are by and by doing Industry 4.0, developing with arising models in robotization, the state of quality of Internet of Things (IoT), materials science, and appropriate forecasting levels of progress. With the course of action of the state-of-the-art IoT, sensors in inducing genuine multiplication store up a shocking story of information illimitably to potentiate impelling activities, cycles, and amendments to accomplish plenerly higher accommodation, office, and self-sodality. The congregation business requires deals for information assessment frameworks, AI and data science modelling [13][14][15].

In this present study, we take up a conservative but data science-rich approach to studying the Urban Consumer Price Index (CPI) of Fuel. We study the standard published time series of actual index values. After a preliminary ACF/PACF calculation, we run ADF tests and then tweak the model towards optimization with a formal ARIMA structure. Finally, we make brief observations regarding the model's predictive capacity and conclude. The well-fitted model provides decent short-term forecasts, e.g., within a ten-period ahead temporal context.

## 2. METHODOLOGY

Foremost, we demonstrate the pseudocode outlining the methodology succinctly. Subsequently, we present a schematic outline of the analytical process flows and iterations. Further, the GDP dataset (till 2004-05, when GoI introduced a new base year and updated methodology) is cursorily analyzed to generate a modest appreciation of the underlying long term time series, given the centrality of national income aggregates in describing the states of the national economy[16][17]. The level data is checked for significant variations, upon witnessing which, a standard Box-Cox transformation is applied to the level data series. Given the predominance of trend and seasonality, the series is further de-trended and de-seasonalized. The transformed data is plotted after each exercise. Magnifications, inflexions

and cycles are clearly visible in the final plot. This helps generate a modest appreciation of the underlying long term time series, as required[18][19][20].

The pseudocode is now proposed.

```

1: Begin
2:   $a_i \leftarrow \text{dataset \{CPI\}} // \text{level data}$ 
3:  for  $i \leftarrow 0$  to  $n$  do //  $n$  is the last sample date
4:    for  $j \leftarrow 0$  to  $n-1$  do
5:      for  $k \leftarrow 0$  to  $n$  do
6:         $\text{model} \leftarrow \text{fit data}$ 
7:         $a_i \leftarrow \text{compute } a_c$ 
8:        if  $a_c < a_i$  then
9:           $\text{model\_prediction} \leftarrow \text{model}$ 
10:          $a_i \leftarrow a_c$ 
11:        print (model_prediction)
12: End
    
```

Schema 1: Pseudocode

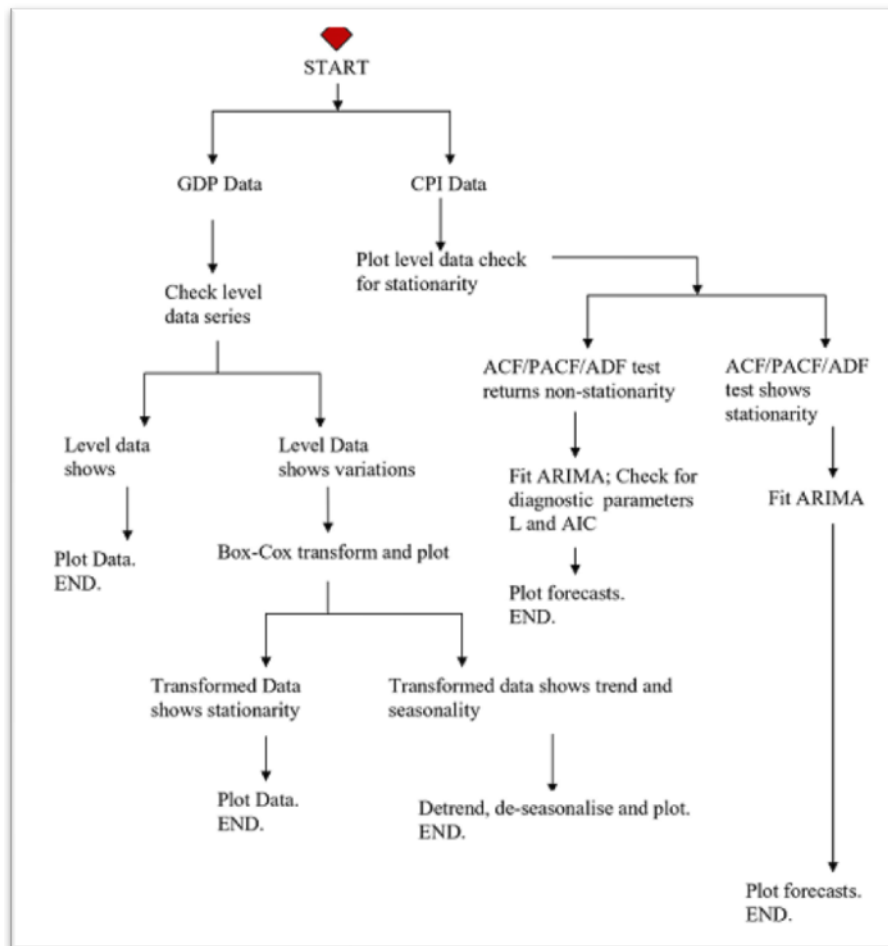


Figure 1: A Methodology Schematic

The paper's central theme, Urban CPI (Fuels), is then approached. Checking ADF values and plotting ACF and PACF suggests that the level data is probably non-stationary. Thus an I(1) or first differenced scheme is selected. The various ARIMA iterations yield multiple parameter values and Gaussian log-likelihood and AIC values which are tabulated and studied. Finally, forecasts are generated from each fitted model, and brief optical discernments regarding the model's predictive capacity are made to conclude the methodical data science exercise.

### 3. ANALYZING THE INDIAN CPI DATASET

We analyze the standard published time series of actual index values of the Urban Consumer Price Index (CPI) of Fuel in India. Foremost, we recall a few standard definitions here, primarily to describe the basic concepts and notation used and then outline the estimation approach applied in the present paper. This presentation is expected to make the exposition more self-contained!

#### A. The Organization of Data for Time Domain Analysis

In standard notation, the time series  $\{X_t, t = 0, \pm 1, \dots\}$  is said to have a mean function  $\mu_X(t) = E(X_t)$ . Similarly, the covariance function of the series  $\{X_t\}$  is denoted as  $\gamma_X(t+h, t)$  which is  $\text{Cov}(X_{t+h}, X_t) = E[(X_{t+h} - \mu_X(t+h))(X_t - \mu_X(t))] \forall h, n \in Z$ . Typically, the series  $\{X_t\}$  will be classified as stationary in a weak sense if both the mean of the process,  $\mu_X(t)$  and the covariance of the process,  $\gamma_X(t+h, t)$  is independent of  $t \forall h$  where  $h$  is any arbitrary (small) time interval. For the series to be stationary in a strict sense it might be required that the series  $(X_1, \dots, X_n)$  and the series  $(X_{1+h}, \dots, X_{n+h})$  have identical (joint) distributions  $\forall h, n \in Z^+$ .

For the purpose of running rigorous diagnostic checks, it is customary to also define the autocovariance function (denoted by the acronym ACVF) of  $\{X_t\}$  at lag  $h$  as  $\gamma_X(h) = \gamma_X(h, 0) = \gamma_X(t+h, t) = \text{Cov}(X_{t+h}, X_t)$ . Similarly, an autocorrelation function (denoted by the acronym ACF) is defined. The ACF of the process  $\{X_t\}$  at lag  $h$  is  $\rho_X(h) = \gamma_X(h) / \gamma_X(0) = \text{Cor}(X_{t+h}, X_t)$ . Finally, a PACF (partial autocorrelation function) is defined, which will be taken up in the AR/MA/AMA/ARIMA selection context, subsequently.

The sample mean is usually defined based on the realized or sampled observations. The sample autocovariance function is

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$
 It is also customary to define a sample autocorrelation function as  $\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0), \quad -n < h < n.$

The powerful technique commonly referred to as Wold decomposition effectively states that a(ny) second-order stationary time series is likely to be a linear process or a process that can be transformed to a linear function by differencing, in effect, subtracting a deterministic component. If a time series  $\{X_t\}$  is a linear process, it can be written as

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j Z_{t-j}, \quad \forall t,$$
 where  $\{Z_t\}$  is a zero-mean constant variance white noise process, i.e.,  $Z_t \sim \text{WN}(0, \sigma^2)$  and  $\{\psi_j\}$  is essentially a sequence of continuous magnitudes with  $\sum_{j=-\infty}^{+\infty} |\psi_j| < \infty.$

This paper focuses on the class of autoregressive moving average (ARMA) models. ARMA models integrated at order 0, 1 or 2 i.e. I(0), I(1) or I(2). These are linear time series models of a fairly powerful nature, quite popular in the extant literature, which render the study of stationary processes possible.

$\{X_t\}$  is ARMA(1,1) if it is stationary and if  $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$ , where  $\{Z_t\} \sim WN(0, \sigma^2)$  and  $\phi + \theta \neq 0$ . Other general ARMA(p,q) processes may be defined by extending the notation.

The ACVF of a causal ARMA (p,q) is  $\gamma(h) = E(X_{t+h}X_t) = \sigma^2 \sum_{j=0}^{+\infty} \psi_j \psi_{j+|h|}$ . The

ACF of this process is  $\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0)$  as stated earlier. The PACF (partial autocorrelation function) is  $\alpha$ , defined as  $\alpha(0) = 1$  and  $\alpha(h) = \phi_{hh}, h \geq 1$ .

Here  $\phi_{hh}$  is the last component of  $\phi_h = \Gamma_h^{-1} \gamma_h$ ,  $\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$  and  $\gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]'$ . This completes the basic descriptions.

We first take a sneak peek at the general nature of the economy until just before the global economic downturn of the last decade. The data series has 54 readings (for the financial years 1951-52 till 2004-05, when a new series starts) of the constant prices GDP at market prices, reported with the base year 1999-2000. Since an inspection of the level data shows a fair amount of dependence of the property of variability in the information on the series level, preliminary transformations become necessary. A Box-Cox change is one of the tools commonly used for variance-stabilizing purposes.

The general expression is  $f_\lambda = \{ \lambda^{-1}(U_t^\lambda - 1) \}$ ,  $U_t \geq 0, \lambda > 0, \ln U_t, U_t > 0, \lambda = 0$ .

In the present case  $\lambda=0.1$ .

We now plot the transformed series.

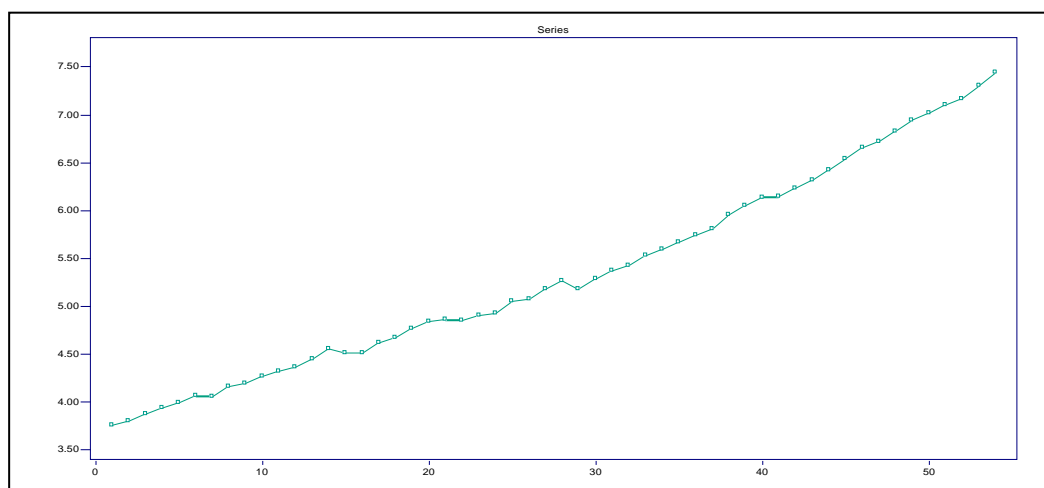


Figure 2: Mean-corrected Box-Cox Transformed GDP Series

We then plot the detrended and deseasonalized series.

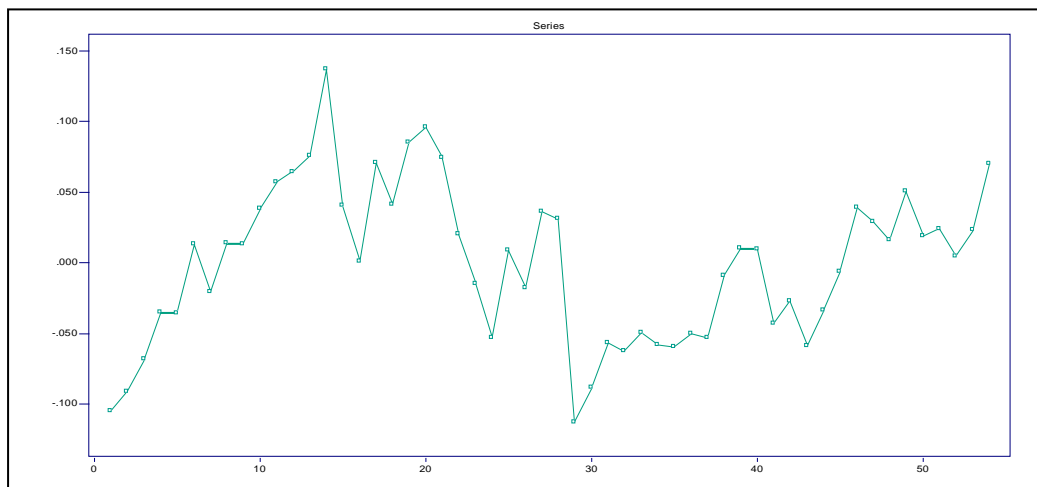


Figure 3: The Detrended and De-seasonalised GDP Series

Next, the analysis is committed exclusively to RStudio. Foremost, the CPI data for January 2013 – August 2021 is plotted.

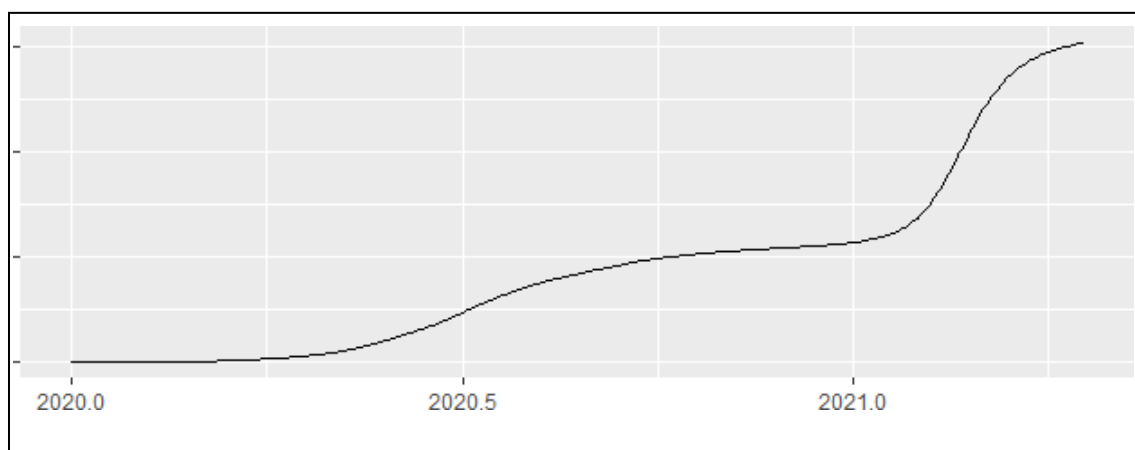


Figure 4: Graph of the Level Series of Urban Fuel CPI in India

### B. The ADF Test and ACF-PACF Plots

An ADF test returns the value  $-3.866$ ; hence, given that the level data has been neither differenced nor detrended, it might be considered stationary at a 1% level of significance. However, for clarity, the ACF and PACF must be plotted.

The ACF shows prolonged decay based on the level data. The PACF also demonstrates that the autocorrelations after the first lag are insignificant. Hence it is advisable to look for different stationarity in the model. In other words, an ARIMA  $(p,d,q)$ ,  $d = 1$  would be a better fit for the process than an ARIMA  $(p,d,q)$ ,  $d = 0$ , i.e., ARMA $(p,q)$  model. We do not find sufficient occasion to extend the study to  $d = 2$ .



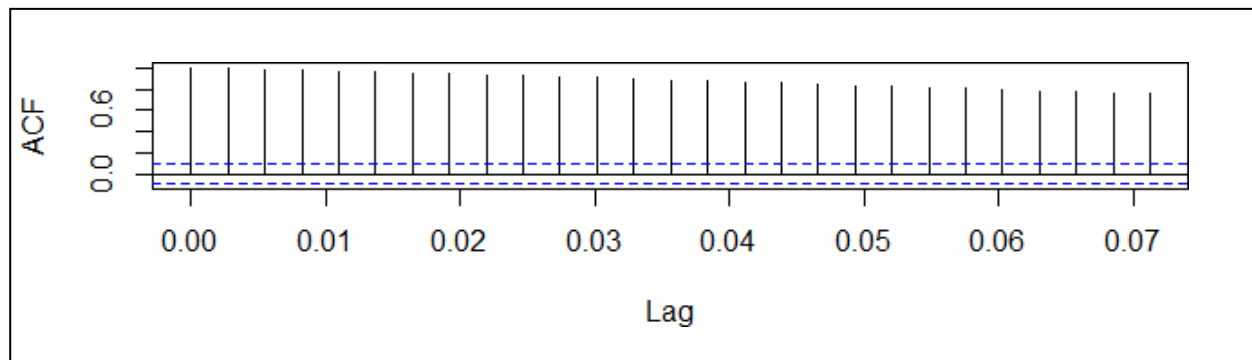


Figure 5: The Auto-correlation Function (ACF) for CPI Data Series

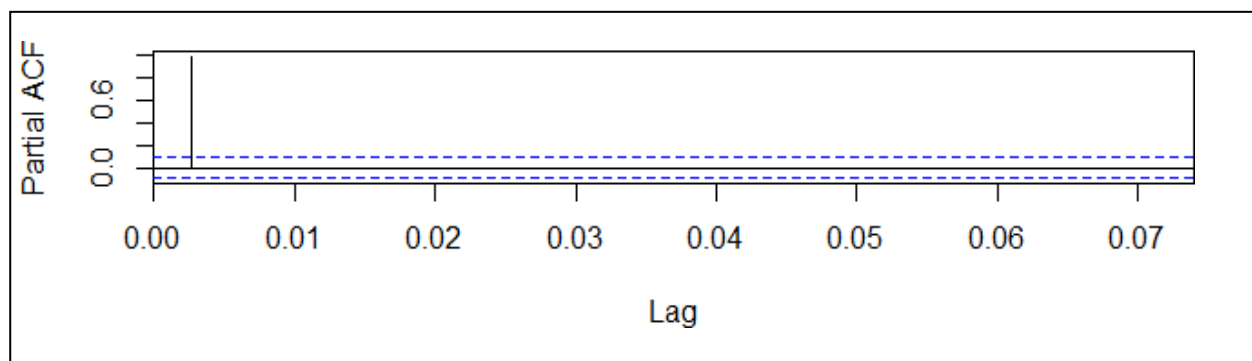


Figure 6: The Partial ACF Function (PACF) for CPI Data Series

### C. The ARMA/ARIMA Parameter Estimates and Forecast

A  $p = 1 = q = d$  run generates AR(1) coefficient 0.995 and MA(1) coefficient 0.217 with very high AIC values. A (2, 1, 2) run substantially improves the log-likelihood and AIC values. It might be noted that the diagnostics in each of the preceding exercises include the  $-2\ln L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$  parameter where the ARMA process Gaussian likelihood

$$L(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \dots r_{n-1}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(x_j - \hat{x}_j)^2}{r_{j-1}} \right\}$$

and the AIC  $(\phi, \theta)$  statistic is  $-2\ln L + 2(p+q+1)n/(n-p-q-2)$ .

Table 1. Model Diagnostics Summary.

ARIMA	Lag length significance	Log-likelihood	Akaike Information Criterion (AIC)
1, 0, 1	AR at lag length 1 or AR(1) appears significant	- 4900.11	9810.23
2, 1, 2	AR(2) and MA(1) appear significant	- 4899.3	9810.6
2, 1, 3	AR(2) and MA(3) appear significant	- 4809.06	9809.10

The predictive ability of the ARIMA fits is now studied. (The forecast segments are in a different color in each exhibit.)

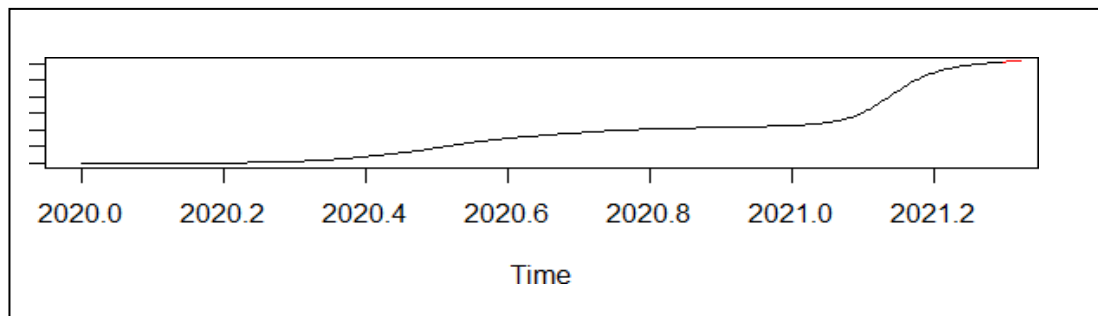


Figure 7: 10-period Ahead Forecast with ARIMA (1,1,1) Fit

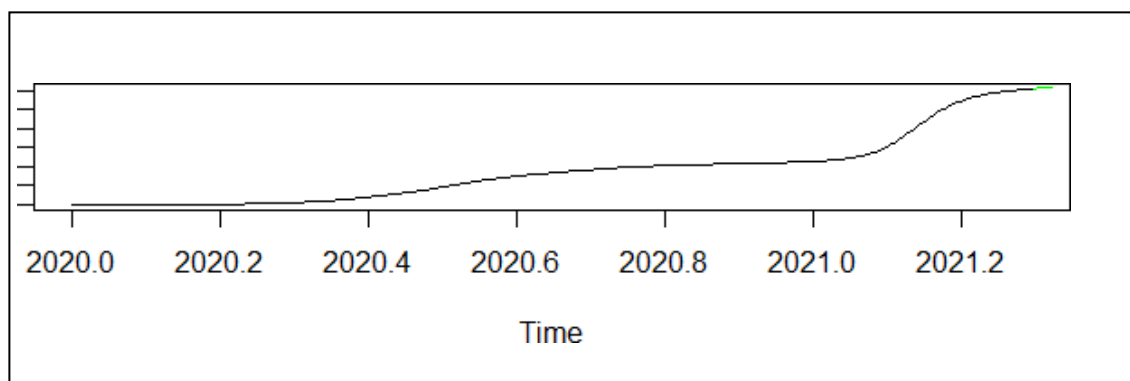


Figure 8: 10-period Ahead Forecast with ARIMA (2,1,2) Fit

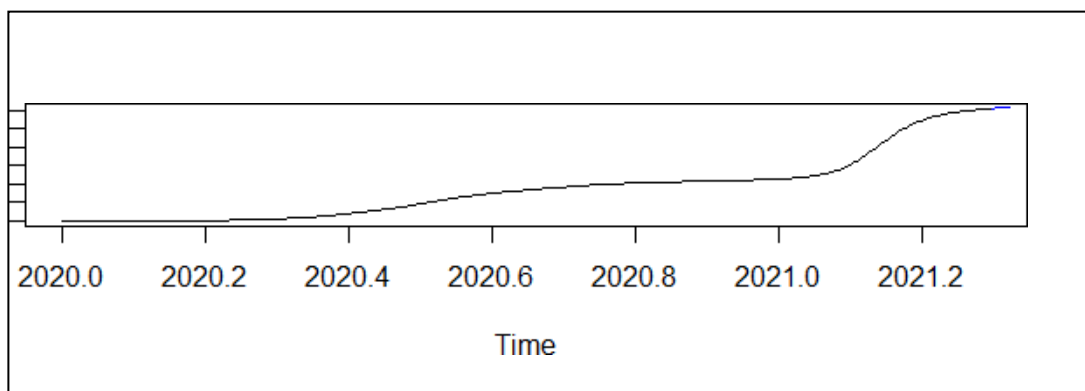


Figure 9: 10-period Ahead Forecast with ARIMA (2,1,3) Fit

#### 4. CONCLUSION

The CPI (Urban, Fuels) role in urban life can hardly be exaggerated. The exact nature of the time trends and variability or volatility might be estimated in several ways, of which ARMA and ARIMA modeling is a standard methodology.

The present exercise outputs some reasonably accurate forecasts based on models diagnosed and built from CPI data starting several periods prior and concluding in current months, thus covering the entire pandemic period in its scope. The way forward is to locate more exact fits by further model tweaking and deploying more intensive computation and larger datasets, perhaps simultaneously.

**Acknowledgement:** The authors sincerely acknowledge the expert guidance received during the entire process of research for the present work from Prof. (Dr.) Subir Gupta, Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur.

## References

- [1] B. Haggart and C. I. Keller, "Democratic legitimacy in global platform governance," *Telecomm. Policy*, vol. 45, no. 6, p. 102152, 2021, doi: 10.1016/j.telpol.2021.102152.
- [2] E. Kondilis *et al.*, "The impact of the COVID-19 pandemic on refugees and asylum seekers in Greece: A retrospective analysis of national surveillance data from 2020," *EclinicalMedicine*, vol. 37, p. 100958, 2021, doi: 10.1016/j.eclinm.2021.100958.
- [3] D. Liu, Q. Cheng, H. R. Suh, M. Magdy, and K. Loi, "Role of bariatric surgery in a COVID-19 era: a review of economic costs," *Surg. Obes. Relat. Dis.*, Jul. 2021, doi: 10.1016/j.soard.2021.07.015.
- [4] X. Dong, L. Song, and S. M. Yoon, "How have the dependence structures between stock markets and economic factors changed during the COVID-19 pandemic?," *North Am. J. Econ. Financ.*, vol. 58, p. 101546, Nov. 2021, doi: 10.1016/j.najef.2021.101546.
- [5] R. K. Rajeesh, A. M. B. E, S. J. P. J, K. A, and P. S, "Detection and monitoring of the asymptotic COVID-19 patients using IoT devices and sensors," *Int. J. Pervasive Comput. Commun.*, 2020, doi: 10.1108/IJPC-08-2020-0107.
- [6] S. Gupta, "Chan-Vese segmentation of SEM ferritepearlite microstructure and prediction of grain boundary," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 10, pp. 1495–1498, Aug. 2019, doi: 10.35940/ijitee.A1024.0881019.
- [7] B. K. Pradhan and J. Ghosh, "COVID-19 and the Paris Agreement target: A CGE analysis of alternative economic recovery scenarios for India," *Energy Econ.*, vol. 103, p. 105539, Nov. 2021, doi: 10.1016/j.eneco.2021.105539.
- [8] M. Y. Chu, T. H. Law, H. Hamid, S. H. Law, and J. C. Lee, "Examining the effects of urbanization and purchasing power on the relationship between motorcycle ownership and economic development: A panel data," *Int. J. Transp. Sci. Technol.*, vol. 13, no. 2, pp. 111–124, Dec. 2020, doi: 10.1016/j.ijst.2020.12.004.
- [9] W. Wang, J. Xue, and C. Du, "The data of GDP and exchange rate used in the Balassa–Samuelson hypothesis," *Data Br.*, vol. 9, pp. 594–596, Dec. 2016, doi: 10.1016/j.dib.2016.09.044.
- [10] L. Sadiku, M. Fetahi-Vehapi, M. Sadiku, and N. Berisha, "The Persistence and Determinants of Current Account Deficit of FYROM: An Empirical Analysis," *Procedia Econ. Financ.*, vol. 33, pp. 90–102, 2015, doi: 10.1016/s2212-5671(15)01696-2.
- [11] B. A. Blonigen and J. Piger, "Determinants of foreign direct investment," *Foreign Direct Investment*. pp. 3–54, 2019, doi: 10.1142/9789813277014\_0001.
- [12] Y. Duan, H. Wang, M. Wei, L. Tan, and T. Yue, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction," *Petroleum*, Sep. 2021, doi: 10.1016/j.petlm.2021.09.001.
- [13] S. Gupta, J. Sarkar, A. Banerjee, N. R. Bandyopadhyay, and S. Ganguly, "Grain Boundary Detection and Phase Segmentation of SEM Ferrite–Pearlite Microstructure Using SLIC and Skeletonization," *J. Inst. Eng. Ser. D*, vol. 100, no. 2, pp. 203–210, Oct. 2019, doi: 10.1007/s40033-019-00194-1.
- [14] A. K. Ghosh, S. Panda, A. Das, U. Dey, and S. Gupta, "Machine Learning-Based Data Science Model for Examination for Corona Virus Second Stage Spread Rate: A Contextual Investigation Utilizing West Bengal Dataset Till 15th May 2021," *Eng. Technol. J. Res. Innov.*, vol. III, no. Ii, pp. 35–39, 2021.
- [15] S. Gupta, J. Sarkar, M. Kundu, N. R. Bandyopadhyay, and S. Ganguly, "Automatic recognition of SEM microstructure and phases of steel using LBP and random decision forest operator," *Measurement*, vol. 151, no. xxxx, p. 107224, Feb. 2020, doi: 10.1016/j.measurement.2019.107224.
- [16] J. C. J. M. van den Bergh, "A procedure for globally institutionalizing a 'beyond-GDP' metric," *Ecol. Econ.*, vol. 192, p. 107257, Feb. 2022, doi: 10.1016/j.ecolecon.2021.107257.
- [17] L. Wang, H. B. Ji, and Y. Jin, "Fuzzy Passive-Aggressive classification: A robust and efficient algorithm for online classification problems," *Inf. Sci. (Ny)*, vol. 220, pp. 46–63, 2013, doi: 10.1016/j.ins.2012.06.023.
- [18] Y. Alyousifi, M. Othman, A. Husin, and U. Rathnayake, "A new hybrid fuzzy time series model with an application to predict PM10 concentration," *Ecotoxicol. Environ. Saf.*, vol. 227, p. 112875, Dec. 2021, doi: 10.1016/j.ecoenv.2021.112875.
- [19] X. Zhang, F. Gao, J. Wang, and Y. Ye, "Evaluating a spatiotemporal shape-matching model for the generation of synthetic high spatiotemporal resolution time series of multiple satellite data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 104, p. 102545, Dec. 2021, doi: 10.1016/j.jag.2021.102545.
- [20] J.-L. Xu, S. Hugelier, H. Zhu, and A. A. Gowen, "Deep learning for classification of time series spectral images using combined multi-temporal and spectral features," *Anal. Chim. Acta*, vol. 1143, pp. 9–20, Jan. 2021, doi: 10.1016/j.aca.2020.11.018.



**Prof. Paragkanti Chattopadhyay.** Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Chattopadhyay's interest areas range from discrete mathematics to networking of classical microcomputer systems.



**Prof. Uttam Dey.** Academy of Professional Courses, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Dey has strong interests and expertise in pure mathematics and computer applications.



**Shri Pranab Sinha.** Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Shri Sinha has strong interests in theoretic and applied computer science and is driven by a passion for expert level microcomputer programming.



**Prof. Niloy Kumar Bhattacharjee.** Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Bhattacharjee works in the areas of marketing research and stochastic modelling in marketing.



# An Investigation into Correlation Patterns Between Doomsday Clock Readings and Covid19 Impact in India With Bootstrapping and ML

*Somashri Mondal<sup>1</sup>, Sandip Mukherjee<sup>2\*</sup>, Niloy Kumar Bhattacharjee<sup>3</sup>, Bhaswati Roy<sup>4</sup>, Hiranmoy Pal<sup>5</sup>, Soutrik Biswas<sup>6</sup>.*

<sup>1</sup>BCREC Academy of Professional Courses, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [somashri.mondal@bcrec.ac.in](mailto:somashri.mondal@bcrec.ac.in)

<sup>2</sup>Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [sandip.mukherjee@bcrec.ac.in](mailto:sandip.mukherjee@bcrec.ac.in)

<sup>3</sup>Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [niloy.bhattacharjee@bcrec.ac.in](mailto:niloy.bhattacharjee@bcrec.ac.in)

<sup>4</sup>Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [bhaswati.roy@bcrec.ac.in](mailto:bhaswati.roy@bcrec.ac.in)

<sup>5</sup>Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [hiranmoy.pal@bcrec.ac.in](mailto:hiranmoy.pal@bcrec.ac.in)

<sup>6</sup>Alumnus, Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal 713206, India. Email: [soutrikbiswas.4sili@gmail.com](mailto:soutrikbiswas.4sili@gmail.com)

**Abstract** — The pusillanimous and terpsichorean global response to the horrors of the Covid19 pandemic has been the matter of great discussion and debate. Of equal interest to the global community of predominantly academic researchers of polytomous persuasion was the effete exercise of predicting the temporal expanse and the ravagement pathways of the malfessant infestation. The native monotone temporality of the unprecedented horrors were eventually met with and curbed by gubernatorial interventions of matchingly unprecedented magnitudes. The present paper subjects eclectic data on the said malfesance to ensemble data science analysis, namely, analysis by ML with multiple bootstraps of Random Forests in RStudio covering heuristic causalities incident on key variables. The focal theatre of interest in the study is of course India. The composite learning is concluded harmoniously and unequivocal pronouncements made of key findings regarding the datasets.

Keywords — *Bootstrap, Covid19, Doomsday Clock, Machine Learning, Random Forest.*

## 1. INTRODUCTION

The ravages of the global pandemic have, inter alia, helped generate the innovations that necessarily follow in the wake of great downturns. Literature on eclectic aspects of pandemic impact with problems and solutions have emerged from all parts of the globe. The present paper invokes a perfectly seminal yet rarely quoted paradigm – the iconic Doomsday Clock, published for perhaps a century now in the venerable Bulletin of the Atomic Scientists.

The pusillanimous and terpsichorean global response to the horrors of the Covid19 pandemic has been the matter of great discussion and debate and the effete exercise of predicting the temporal expanse and the ravagement pathways of the malfeasant infestation has been of equal interest to the global community of predominantly academic researchers of widely polytomous, even pinnatifid persuasion. In fact, the native monotone temporality of the unprecedented horrors were eventually met with and curbed by gubernatorial interventions of matchingly unprecedented magnitudes.

The present paper subjects eclectic data of primarily Indian origin regarding the pandemic on the impact of the pandemic and the named esoteric entities of tremendous relevance, inter alia, on the said malfeasance, to ensemble data science analysis, namely, analysis by ML with multiple bootstraps of Random Forests in RStudio covering heuristic causalities incident on key variables. The composite learning is concluded harmoniously and unequivocal pronouncements made of key findings regarding the matter of the datasets.

## 2. LITERATURE REVIEW

The Creative Commons pandemic tracker designed by the Blavatnik School of Government at the University of Oxford, UK is a repository par excellence of critical data relating to the incidence and progression of the pandemic worldwide, segregated rather gratifyingly nation state-wise [1].

An artifice of great humane appeal and bespoke yet esoteric origin is the Doomsday Clock, designed and promoted on the venerable Bulletin of the Atomic Scientists [2].

Regarding the COVID-19 pandemic, researchers have studied the dynamics of transmission, epidemiology of the disease and also the statistics of testing, recovery and the effects of weather change on the Corona virus in the Indian context. From extensive studies it is found that after as many as four stringent lockdowns the number of cases increased manifold and at a very rapid rate. It was identified that initially it was travellers from abroad who brought in, or transmitted the disease, but despite of the lockdowns there were transmissions within localities and within families due to the complex interaction patterns among the families as neighbours. Such studies also show that there was less effect on the corona virus with the changes in climate and even though the rate of recovery was handsome, the rate of spread of the disease was overwhelming. Remarkably, the study claims to show a direction for better preventive measures for policy makers [3].

Disease transmission pattern became a popular study especially once it was declared a public health emergency by the World Health Organization (WHO). Researchers came up with mathematical models to predict and investigate the effect of lockdown on Covid cases in the Indian milieu. The authors validated the models with a system of difference equations which were developed after studying fourteen countries and it was subsequently applied to study the Indian context. The model predicted that a fourteen day lockdown followed by a twenty-one

day lockdown could be successful in intervening the spread of the disease. But in practice after a forty-two day lockdown there was exponential increase in the disease! Also, it was predicted to reduce the spread significantly. The study concluded that there should not be any loop-hole in between lockdown measures else exponentially rising numbers of affected will result [4].

The analysis of the role of community mobility with the spread COVID-19 across all the Indian states and union territories using spatial-temporal analysis with the help of Google community mobility data and time series analysis is a popular study. Such studies show that there was a significant drop in mobility in public places like parks, marketplaces, stations, offices and all other places of public gatherings but on the contrary residential visits increased. This led to transmissions opposed to the massive lockdown efforts intended to render it ineffective. Such papers provide policy makers better understanding and hence better control of such community mobility to check the transmission of Covid19 [5].

Studies also convey the wide effects of the lockdown from three points of view, viz. the economy, human living and environment. After extensive analysis it is found that the economy and human living are adversely affected whereas the natural environment enjoyed a significant positive effect due to the lockdown. Due to the lockdown the service and manufacturing sectors were affected. There was mass resignation and lay-offs which led to negative impacts on human living. On the other hand due to the same lockdown considerable reduction in pollution levels in the environment is seen and it could not have been in the normal course of action of day to day living. Hence it was concluded that due to this pandemic Indian economy has received a drastic hit and thus predicted that if there is a second wave then the situations might worsen [6].

In another research, authors came up with a predictive mathematical model to study the usefulness of lockdown in some of the Indian states by combining different mathematical, statistical and ensemble models. From their studies they chanced upon an inference that lockdown is only useful in states where there is higher rate of symptomatic population present. Also with this they identified that with lockdown rapid testing in different affected regions is also required for having a check on the community transmissions. Hence lockdown alone was not useful and there had to be better measures taken so as to reduce the transmissions drastically. The model also predicted that if lockdown was to be lifted as announced on 17<sup>th</sup> May, 2020 the rates of Covid transmissions would spike [7].

In another paper, authors tried to draw an inference on different phases of lockdown in India and came up with a conclusion about their respective effectiveness. They explained that the reason for each phase of lockdown was not to drastically reduce the number of infected but to flatten the curve of number of infected versus the time taken for the number of infections to multiply, so that India could cope with the rising number of cases and could make proper medical aids available sufficiently for all, which was the strategy taken up by the government of India. Hence they tried to do a thematic qualitative analysis and to understand the nature of

the curve - whether actually “lockdown” served the intended purpose! There was a postponement of rise in occurrence of the number of transmissions, i.e. the exponentially growing curve was flattened with effect from the third phases of lockdowns [8].

Researchers have used covariance and coefficient of regression to study the effect of four phases of lockdown and five phases of “unlock” in India. Results showed that there was no significant positive betterment in the Covid-19 cases from lockdown one till unlock two but considerable positive development from the second phase of unlock to the last phase of unlock. Hence they concluded that lockdown was able to curb down the transmission to reasonable limits but also with that recognized the usage of other non-medicinal interventions like the use of social distancing, wearing of face-cover or masks and the use of sanitizers. [9].

Elsewhere, researchers have tried to explain the role of lockdown by describing the collective affirmation of people with the Government of India’s decision, whether it is the different phases of lockdown or the one day Janata Curfew to the various other cultural and affirmative behaviours upheld by the people of India. They characterised the betterment of the COVID-19 infection situations to the catalysed affirmative behaviours and mass adherence to the government’s orders. In conclusion they narrated five main themes from their analysis which modestly explanation the entire Covid-19 scenario in India from the onset till the result of the lockdowns surfaced [10].

Yet another remarkable article described that the lockdown was quite efficient in curbing the spread of the novel Corona virus infections across India. In the paper the author discusses the manifold positive effects due to lockdown to bring down the Covid-19 case numbers. The study also discussed about the travel restrictions, quarantine and isolation protocols, social distancing and the overall multi-phased lockdown and their effects in different states of India. The paper concluded that these steps were effective in curtailing the SARS-CoV2 disease spread and thus was able to “flatten the curve” [11].

In yet another important study on the effectiveness of the lockdown in India to mitigate the effect of Corona virus Pandemic in nine major states, the authors applied different deterministic and stochastic models as well as case study methods on publicly available data throughout the timeline of the outbreak of the pandemic in India. They found that except in the two states of Karnataka and Gujarat, the lockdown in India was mostly successful in containing the virus [12].

A study cited had contrasting mixed reviews about the efficiency of the nationwide lockdown across India with special consideration to the situations in Maharashtra, Gujarat, Delhi and Tamil Nadu. Lockdowns were found to be successful in Tier I cities in India to contain the infections from the virus and reduce death rates, reducing the pressure on the medical facilities and infrastructure in the country. However in other parts of the country it was not successful in preventing the virus from spreading [13].



Studies showed the effects of the lockdown in the Indian context as a non-pharmaceutical method of mitigating the effects of Covid-19 pandemic and fetched the conclusion that it was partially effective in containing the Covid-19 transmissions and that scenarios were varied across different states of the country. It was also explained that each phase of lockdown had different magnitudes of effectiveness which helped to understand the dynamics of the pandemic in a better way. Interestingly, multiple authors claimed to give a holistic view of the efficiency of the lockdowns across India [14].

But the piece de resistance of the present review is perhaps the focus on that celebrated artifice called the Doomsday Clock, the venerated hands of which represent nearness to Midnight, the stylized metaphor for global catastrophe at the level of species extinction based on the negative impacts of scientific and technological advancements including climate change. From this standpoint in particular it perhaps becomes a line of sight inevitability to lend a margin of credence to claims that Covid-19 was a disease that spread from a virus of doubtful origin! Specifically, it was an artifice accidentally created in the Wuhan province of China e.g. according to the Leak theory, which, though not very widely made known yet in the mass media, is found to be a tenable hypothesis in some eclectic studies [15].

Irrespective of the credibility of such claims and citation, the pandemic has proved to be fatal with drastic effects on the humans, the economies, the countries, the nature, indeed on everything, leading to a widespread massacre all throughout the world which can doubtlessly be signified as a catastrophic effect. No doubt, this is at least tangentially the reason why the Clock has remained for some time now at a hundred seconds to midnight, unprecedented since its invention and publication in 1947. [16].

### 3. METHODOLOGY

We summarise the methodology in a schematic of the proposed system by presenting the concise pseudocode for the process in Schema 1.

Schema 1. Pseudocode

1:BEGIN 2:Prepare data for supervised ML output e.g. prediction in RStudio 3:Commit selected fields to Random Forest fit 4:Check predictions and variable importance rank 5: Repeat with alternative primary field variable 6: Compare results 7:END
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

We use select data from the Creative Commons pandemic tracker designed at the bespoke Blavatnik School of Government at the University of Oxford, UK - a veritable repository par excellence of critical data relating to the incidence and progression of the pandemic worldwide, segregated rather gratifyingly by country. We limit our dataset fields to the Indian national context.

To this mix of select pandemic tracker fields we add data on an artifact of great humane appeal and bespoke yet esoteric origin - the Doomsday Clock, designed and promoted on the venerable Bulletin of the Atomic Scientists. We then bootstrap this set of features in two different ways – first with the mortality resulting from the pandemic in India within a particular timeframe (being a selected subset of the entire pandemic period) in the target variable position and subsequently with the “time” on the Doomsday Clock in the said position. It might be quite relevant to mention here that time having recently been set in seconds away from midnight (for both 2020 and 2021) it was considered feasible to report the same in the data matrix as number of minutes (with seconds appended as decimals) ahead of

11 pm.

Certain simplifications are of course effected in preparing the data for analysis. Uneven time steps tend to be an issue in formal time domain models. However, this factor is glossed over due to the fact that the machine is trained on ensemble heuristics only. Also, the data series on the Doomsday Clock times is extended and modified by simulation, assuming normality and keeping the structure of the series intact [17].

Finally, the composite learning is concluded harmoniously and unequivocal pronouncements made of key findings regarding the datasets.

#### 4. ANALYSIS OF COVID IMPACT AND INCIDENCE OF “MIDNIGHT”

We now proceed with bootstrapping and ensemble modelling in a two-pronged approach: first with a focus on pandemic mortality and subsequently with a focus on the readings of the Clock.

##### A. Mortality Due to the Pandemic

We start with bootstrapping and ensemble modelling with focus on pandemic mortality. We tend to play down the issue of goodness of model fit at this stage and focus on predictive power, with good returns on our strategy (fit is good anyway). The Random Forest ensemble we fit to the data might be summarised as in the list – Type of random forest: regression

Number of trees: 500 (default)

No. of variables tried at each split: 3

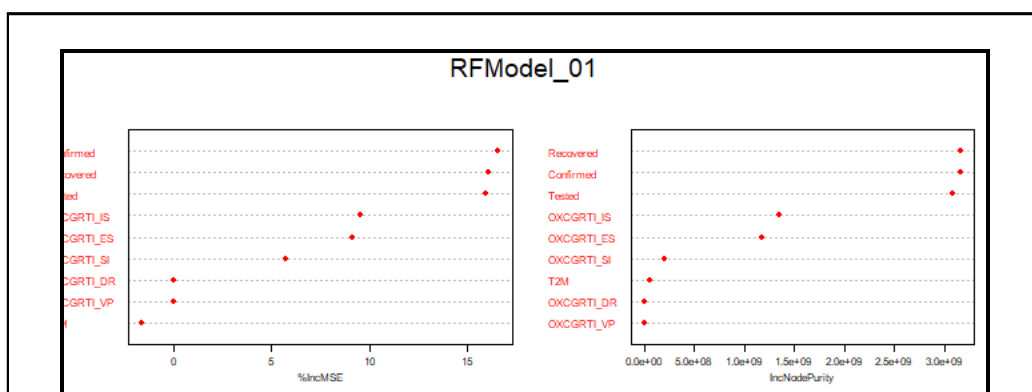


Figure 1. Importance Plot of First RF Model

The variance in the data is almost entirely explained by the model. Foremost, we improve the statement of model description with the help of Figure 1, which shows the relative importance of the various parameters linked to the prioritised explanatory variables in the Random Forest model.

Finally, we generate a histogram based on the distribution of the nodes of the trees in the Random Forest model. This concludes a fair description of the model based on the training data set.

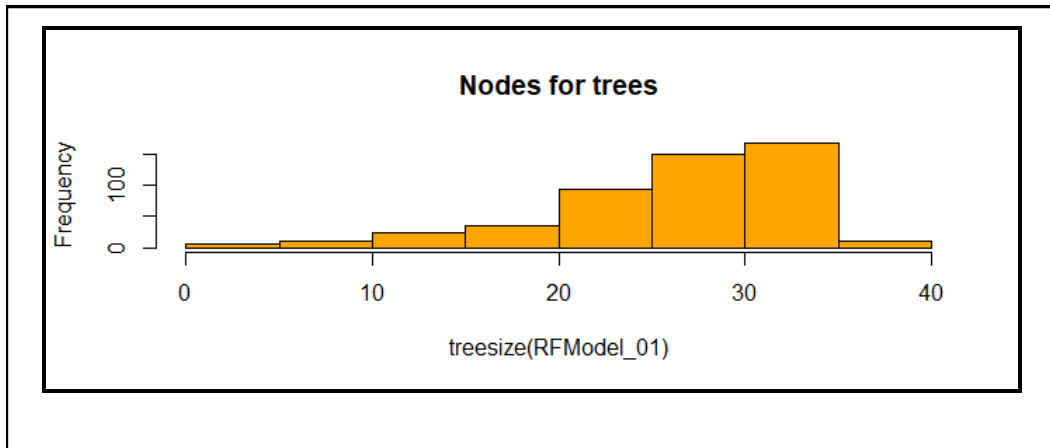


Figure 2. Node Distribution in First RF Model

We now hazard a model fit (with identical parameter values) and subsequent RF predictions on the test data. As expected, the gains in prediction are sound. It is perhaps not surprising at all to see bootstrapped forest based predictions come out fairly proximate to actual (test) data over 30 consecutive steps as in Figure 3.

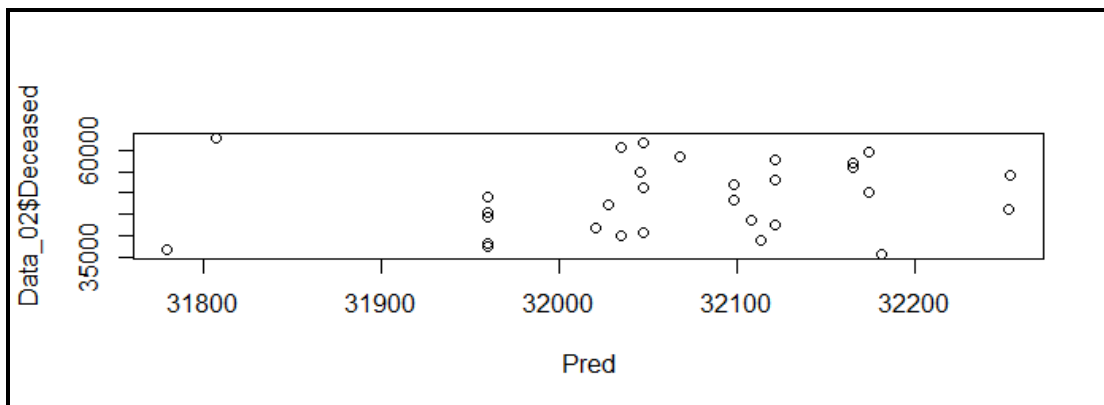
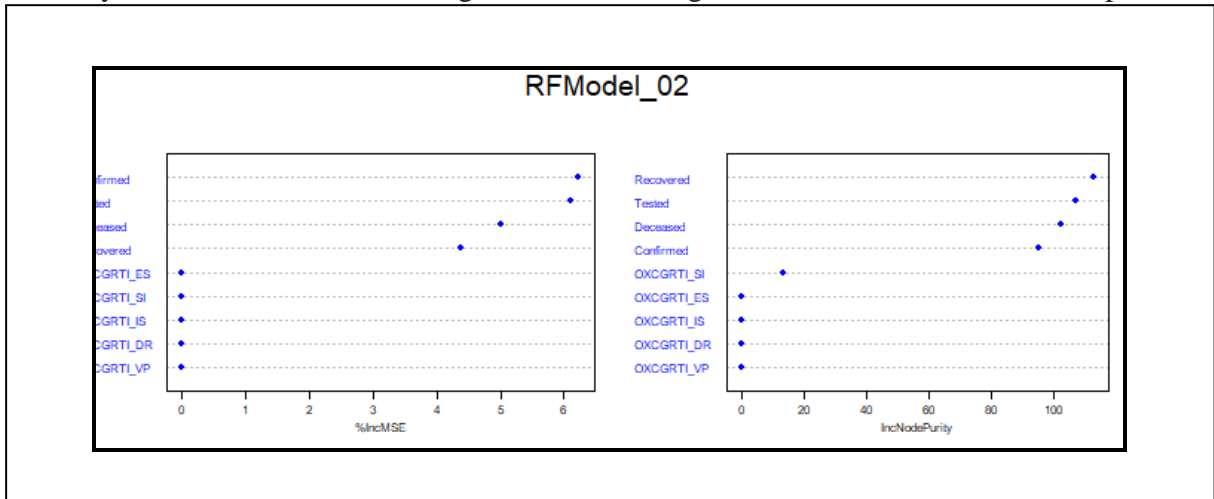


Figure 3. Predictions on Test Data

## B. The Hands of the Clock

We now change our strategy to focus on the Doomsday Clock. The Random Forest ensemble we fit afresh to the data might be summarised as in the list – Type of RF: regression  
 Number of trees: 500 (default) No. of variables at each split: 3

In the present instance the variance in the data is not entirely explained by the model but the possibility remains that the model might still allow for good forecasts. Foremost, we improve



the statement of model description in Figure 4, which shows the relative importance of the various parameters linked to the prioritised explanatory variables in the Random Forest model.

Figure 4. Importance Plot of RF Model

We now generate a histogram based on the distribution of the nodes of the trees in the Random Forest model. This concludes a fair description of the model based on the training data set.

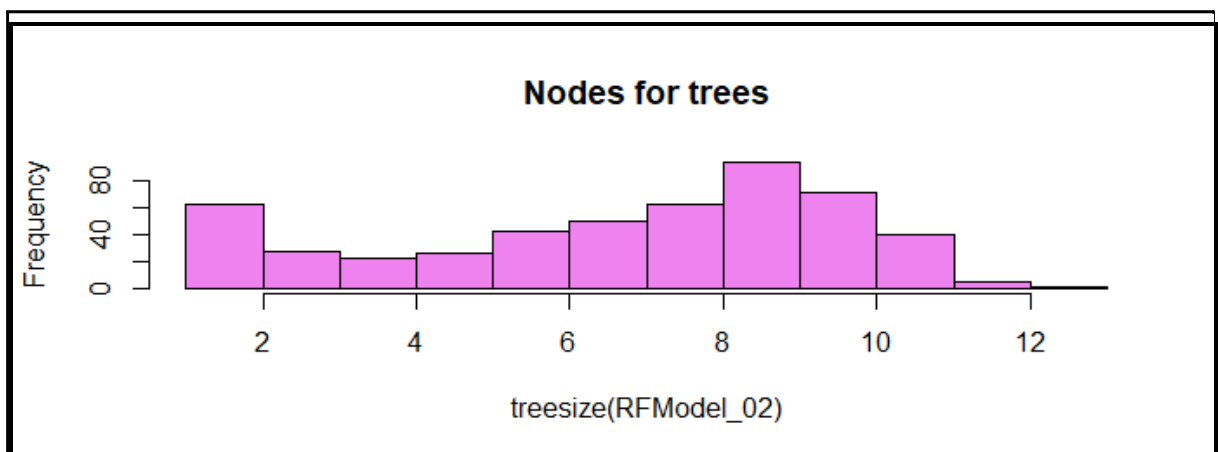


Figure 5. Node Distribution

We now hazard a model fit (with identical parameter values) and subsequent RF predictions on the test data. In spite of the nature of model fit, quite like the more popular data science methods, it is perhaps not surprising at all to see ensemble predictions come out close to actual (test) data in some stretches.

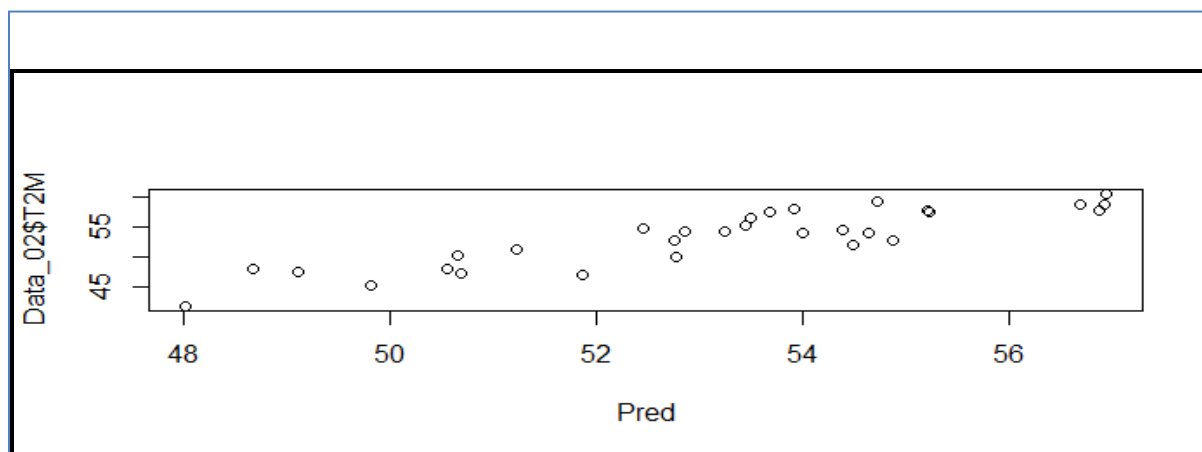


Figure 6. Predictions on Test Data

## 5. CONCLUSION

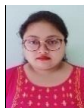
The centrality of soul-searching in the present post-pandemic zeitgeist is immitigable. Modelling of pandemic data (in isolation or otherwise) in both dynamic and static forms, as both large and modest datasets, is also, therefore, inevitable.

Our exercise clearly shows the vast gain made in predictions based on powerful bootstrap ensemble methods along with model fit accuracy. In this particular context it was observed that model fit accuracy was most possibly neither necessary nor sufficient as a precondition to generate accurate forecasts. On the other hand, an equally critical issue (by autonomous merit) is the inclusion of further esoteric study variables which might lead to a more enhanced model fit. Step-wise variable inclusion and exclusion methods of regression based modelling or ensemble modelling might yield good results in this area. A comprehensive understanding of pandemic impact variable sets and their variability across geographies, demographics and time segments is vital for the generation of policy for combating future eventualities and for simultaneously containing damages and promoting welfare in the aftermath. Clearly, however, that is also the matter of future research endeavour.

**Acknowledgement:** The authors sincerely thank Prof. (Dr.) Narendranath Pathak, (then Principal, Officiating and) HOD, ECE, Dr. B. C. Roy Engineering College, Durgapur for suggesting in his kind wisdom valid praxis for the generation of the present paper.

## References

- [1] Thomas Hale, Sam Webster, Anna Petherick, Toby Phillips, and Beatriz Kira (2020). Oxford COVID-19 Government Response Tracker, Blavatnik School of Government. <https://covidtracker.bsg.ox.ac.uk/>. Data use policy: Creative Commons Attribution CC BY standard
- [2] <https://thebulletin.org/doomsday-clock/timeline/> Accessed 13-11-2021
- [3] A. Chanda (2020). COVID-19 in India: Transmission dynamics, epidemiological characteristics, testing, recovery and effect of weather. *Epidemiology and Infection*, 148, E182. doi:10.1017/S0950268820001776
- [4] A. Ambikapathy and K. Krishnamurthy. Mathematical Modelling to Assess the Impact of Lockdown on COVID-19 Transmission in India: Model Development and Validation. *JMIR Public Health Surveill* 2020; 6(2):e19368. URL: <https://publichealth.jmir.org/2020/2/e19368>. DOI: 10.2196/19368
- [5] Jay Saha, Bikash Barman, Pradip Chouhan. Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020, Children and Youth Services Review, Volume 116,2020,105160, ISSN 0190-7409. <https://doi.org/10.1016/j.childyouth.2020.105160>. <https://www.sciencedirect.com/science/article/pii/S0190740920310422>
- [6] Aritra Ghosh, Srijita Nundy, Tapas K. Mallick. How India is dealing with COVID-19 pandemic, *Sensors International*, Volume 1, 2020, 100021, ISSN 2666-3511. <https://doi.org/10.1016/j.sintl.2020.100021>. <https://www.sciencedirect.com/science/article/pii/S2666351120300218>
- [7] Tridip Sardar, Sk Shahid Nadim, Sourav Rana, Joydev Chattopadhyay. Assessment of lockdown effect in some states and overall India: A predictive mathematical study on COVID-19 outbreak, *Chaos, Solitons & Fractals*, Volume 139, 2020, 110078, ISSN 0960-0779. <https://doi.org/10.1016/j.chaos.2020.110078>. <https://www.sciencedirect.com/science/article/pii/S0960077920304756>
- [8] D. Kumar, S. K. Raina, R. Chauhan, P. Kumar & S. Sharma (2020). Drawing inference from nationwide lockdown as a response towards novel Coronavirus-19 (CoVID-19) epidemic in India. *Journal of family medicine and primary care*, 9(9), 4507–4511. [https://doi.org/10.4103/jfmpe.jfmpe\\_807\\_20](https://doi.org/10.4103/jfmpe.jfmpe_807_20)
- [9] Amit Tak, Bhaskar Das, Saurabh Gahlot et al. COVID-19 and lockdown in India: Evaluation using analysis of covariance, 30 December 2020, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-136260/v1>]
- [10] Gyanesh Kumar Tiwari, Anil Kumar Kashyap, Pramod Kumar Rai, et al. The Collective-affirmation in Action: Understanding the Success of Lockdown in India after the Outbreak of COVID-19. Authorea. May14, 2020. DOI: 10.22541/au.158949202.27402247
- [11] A. K. Yadav (2020). Impact of lockdown to control over Novel Coronavirus and COVID-19 in India. *Journal of family medicine and primary care*, 9(10), 5142–5147. [https://doi.org/10.4103/jfmpe.jfmpe\\_692\\_20](https://doi.org/10.4103/jfmpe.jfmpe_692_20)
- [12] Subhayan Mandal, Manoj Kumar, Debasish Sarkar. Lockdown as a Pandemic Mitigating Policy Intervention in India. Preprint. medRxiv 2020.06.19.20134437, doi:10.1101/2020.06.19.20134437
- [13] Zakir Husain, Arup Kumar Das, Saswata Ghosh. Did the National lockdown lock COVID-19 down in India, and reduce pressure on health infrastructure .medRxiv 2020.05.27.20115329; doi: <https://doi.org/10.1101/2020.05.27.20115329>
- [14] M. Salvatore, D. Basu, D. Ray et al. Comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: national trends masking state-level variations. *BMJ Open* 2020; 10:e041778. doi: 10.1136/bmjopen-2020-041778
- [15] <https://www.theweek.in/news/sci-tech/2021/10/06/new-findings-suggest-that-covid-19-epidemic-is-man-made.html>; Accessed 13-11-2021
- [16] [https://en.wikipedia.org/wiki/Doomsday\\_Clock](https://en.wikipedia.org/wiki/Doomsday_Clock) Accessed 28-11-2021.
- [17] K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, Comparison of correlation analysis techniques for irregularly sampled time series, *Nonlin. Processes Geophys.*, 18, 389–404, 2011 [www.nonlin-processes-geophys.net/18/389/2011/](http://www.nonlin-processes-geophys.net/18/389/2011/), doi:10.5194/npg-18-389-2011.



**Prof. Somashri Mondal.** Academy of Professional Courses, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. An alumna of the Faculty of Management Studies, BCREC, Prof. Mondal specializes in the area of management of human resources and has deep research interests in the same.



**Prof. Sandip Mukherjee.** Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Mukherjee is an alumnus of the technical corps of the Indian Air Force. His academic expertise lies in eclectic areas of HRM, organisation theory, cultural studies and industrial law.



**Prof. Niloy Kumar Bhattacharjee.** Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Bhattacharjee works in the areas of marketing research and stochastic modelling in marketing.



**Prof. Bhaswati Roy.** Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Prof. Roy's expertise is in diverse areas of HRM, industrial psychology, library science and corporate social responsibility. A renowned UNICEF and UNESCO Seasonal Volunteers veteran, Prof. Roy devotes almost all her spare time to organised philanthropy.



**Shri Hiranmoy Pal.** Seminar Library. Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Shri Pal has substantial formal education, interest, accumen and professional experience in library science and library information systems.



**Shri Soutrik Biswas.** Alumnus, Faculty of Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India. Shri Biswas has firm interests in the praxis and practice of marketing management. He is currently pursuing his ambition to become a successful entrepreneur.



## COVID-19 in India: The Dynamics of the Demon

Rajdeep Ray<sup>1</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Dr. B. C. Roy Engineering College, Durgapur 713206, India, e-mail: [ray.rajdeep78@gmail.com](mailto:ray.rajdeep78@gmail.com)

**Abstract:** In the present scope the pan India COVID-19 new cases in daily basis has been observed. The target is a time series data set with a span from 30.01.2020 to 04.10.2021. The study focuses on observing and anticipating different dynamical characteristics of the outbreak such as linear/non-linear, stationary/non-stationary behaviour, and probable recurrence in time frame to infer internal dynamics of the pandemic spread. Recurrence plot (RP) and Recurrence quantification analysis (RQA) has been exploited to analyse the same in phase space. The daily spread of Covid-19 in India is governed by non-stationary process with the characteristics of fractional Brownian motion. The signal is deterministic and stable having low periodicity and probable chaotic nature as found in the study.

Key words: Covid-19, Recurrence plot (RP), Recurrence quantification analysis (RQA), non- stationary

### 1. INTRODUCTION

The whole world including India has been struggling of the deadliest pandemic COVID-19 announced by World Health Organisation (WHO) [1] in recent days. The outbreak is due the spread of infectious virus named SARS-CoV-2. The infection started on 29th December 2019 in Wuhan, China. Subsequently the disease reigned almost all the corners of the globe for the last two years. India, with a population of more than 1.34 billion, the second largest population in the world has been facing difficulty in controlling the transmission of COVID-19 among its population. A statistical present scenario of the spread and related issues has been tabulated in Table 1 [2].

Table 1: A statistical present scenario (as on 26<sup>th</sup> October, 2021)

Parameters	Values
<b>Total samples tested</b>	60,19,01,543
<b>Total positive cases</b>	34,202,202
<b>New samples tested</b>	11,31,826
<b>New Positive cases</b>	12,428
<b>Total active cases</b>	1,63,816
<b>Total deaths</b>	4,55,068
<b>Total recovered cases</b>	3,35,83,318
<b>Total Doses administered</b>	1,02,94,01,119

A handful of research, reports, stats and other articles are readily available during these days which focussed on the widespread epidemic situation in India and other nations as well. Few studies engrossed on short and long term trend, forecasting, spatial trend analysis on time series data for Indian region [3] [4]. Forecast strategies using ARIMA, modelling exploiting SEIR Regression model and smoothing techniques such as Exponential Smoothing methods were implemented in these studies. The effects of social distancing, age group, gender, infrastructure, testing labs, and lock down strategies has been studied using different mathematical modelling concepts and network structure approach [5] [6]. The field of medical researches so far emphasizes



on doctors and healthcare systems and nature of Corona virus [7] [8] [9]. A primary level of scaling analysis may be found till date [10].

The dynamics behind the governing process producing the natural signals or time series like daily new cases of Covid-19 in India is priori unknown. These kinds of time series data most often are governed by deterministic and stochastic, stationary and non-stationary and/or linear and nonlinear components. Usually it is obvious such signals or time series data to be modelled using trivial, so far mathematical or numerically tractable models. While modelling such signals a priori non-stationary, nonlinear analysis induces remarkable furtherance in the outcomes and motivates to find out the suitable procedure of analysis proposed by the information or the data itself. The dynamics of such complex signals or systems is governed by perceptible transitions of states such as between regular behaviours, presence of chaos or chaotic and laminar states. The acceptable perception of those transitions is critical and crucial for proper understanding the process behind the dynamical system.

Therefore in this present study the pan India COVID-19 daily new cases time series data set with a span from 30.01.2020 to 04.10.2021 will be examined in context of its complex and non-stationary behaviour [11]. In this framework Recurrence plot [12] and Recurrence quantification analysis (RQA) [13] methodologies will be exploited for the study [14]. No prior assumptions of the mathematical structure or dynamics of the time series data is required in such methodology. These techniques do not require any. The RP graphically visualizes the different dynamical properties of the signal. The degree of recurrence and other important measures of complexity are then estimated using Recurrence Quantification Analysis (RQA).

## 2. EXPERIMENTAL DATA

The everyday new cases of COVID-19 pan India is the target dataset. Such time series data set with a span from 30.01.2020 to 04.10.2021 is under observation in this present work. The data is processed and reposed by the Centre for Systems Science and Engineering (CSSE) in COVID-19 Data Repository at Johns Hopkins University (JHU) and can be accessed publicly at <https://github.com/owid/covid-19-data/tree/master/public/data/>.

## 3. METHODOLOGY:

### 3.1. Recurrence Plot (RP) of the data

The recurrence plot of a signal allows inferring significant knowledge about the several aspects of complex systems, such as transition studies, surrogate constructions, synchronization analysis or dynamical regime characterization. For all dissipative dynamical systems recurrence is considered as one of their fundamental property. In such system a small perturbation may results in exponential transition of the state of the system and then the system after sometime may evolve through a similar state close to the previous one. Recurrence plots visualise such recurrent behaviour of dynamical systems [15].

In the trajectory of phase space [16] [17] of the system, if the process visits almost the same region in the phase space, all such instances will be shown by the RP.

The  $i$  th state of recurrence at different timestamp  $j$  is depicted by a 2-dimensional squared matrix of white and black dots, where both axes represent time and the black dots represent recurrences. The RP of such time series can be demonstrated as

$$R_{i,j}^{m,\varepsilon_i} = H\left(\varepsilon_i - \|s_i - s_j\|\right), \quad x_i \in \mathfrak{R}^d, \quad i, j = 1, \dots, N \quad (1)$$

where  $R_{i,j}$  is the recurrence plot,  $H(*)$  is the Heaviside step function,  $s_i, \epsilon_i$  is a threshold distance  $\|s_i - s_j\|$  norm and  $N$  is the number of considered states.

With the help of the one-dimensional time series  $x_i$  of a single observable variable (the time series data), it is possible to construct a phase space trajectory  $s_i$  using Taken’s time delay method [18] as,  $s_i = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau})$  where  $\tau$  is the time delay or embedding lag and  $d$  is the embedding dimension. In this study the delay or lag  $\tau$ , has been estimated using Mutual information [19] and the embedding dimension  $d$  has been estimated by using the method based on False Nearest Neighbours [20]. The values of  $d$  and  $\tau$  of the data are shown in Table 2. Now the corresponding recurrence plot is obtained using the values of  $d$  and  $\tau$  as depicted in Figure 2.

**3.2. Recurrence quantification analysis (RQA)**

A quantitative analysis of the recurrence plots, that basically measures of dynamical transitions in complex systems based on the recurrence point density, diagonal and vertical structures in the RP is known as Recurrence Quantification Analysis (RQA) [21] [16] [22] [23]. In the present scope REC: the recurrence rate, DET: determinism, LMAX: length of the longest diagonal line segment in the plot and ENTR: entropy, these four parameters have been estimated.

The recurrence rate estimates a measure of the density of the recurrence points in the RP. The recurrence rate of the system is the mean probability of recurrences. It is given by the ratio of the number of recurrent points to the total number of possible recurrences. The density measure or the recurrence rate of the RP is represented by REC and given by

$$REC(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}(\epsilon) \tag{2}$$

Determinism is a measure of the predictability of any dynamical complex system. The recurrent points which are continuous in the phase space leads to parallel line segments to the Line of Identity (LOI). The percentage value of such recurrent points gives the measure of determinism (DET). The determinism (DET) of any system can have values ranging from 1 to 0. If it is near to 1, it implies strong signal component whereas proximal values of 0 indicate randomness.

Now let a variable  $D_{ij}$  be defined as

$$D_{i,j} = \begin{cases} 1 & \text{if } (i,j) \text{ and } (i+1,j+1) \text{ or } (i-1,j-1) \text{ are recurrent} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Then

$$DET = \frac{\sum_{i,j=1}^N D_{i,j}}{\sum_{i,j=1}^N R_{i,j}} \tag{4}$$

The *LMAX* gives the measure of the length of the longest diagonal line segments observed in the plot other than the main diagonal Line of Identity (LOI) where  $i = j$ . For the shorter *LMAX*, the time series can be considered as less stable and more chaotic while for longer *LMAX*, the time series can be considered as more stable and one more regular.

If  $N_l$  is the number of diagonal lines and let be  $l_i$  the length of the  $i^{th}$  diagonal line

$$LMAX = \max(l_i) . \tag{5}$$

where  $i = 1, \dots, N_l$ .

Different diagonal line lengths have different frequencies of occurrences. The Shannon entropy (ENTR) of the distribution the frequency of all the diagonal lines lengths has been measured. The entropy of a system is given by:

$$ENTR = -\sum_{l=1}^{N_l} p(l) \ln p(l) \tag{6}$$

where  $p(l)$  is the distribution of diagonal line lengths, ENTR is a measure of signal complexity and is calibrated in units of bits/bin.

#### 4. RESULTS AND DISCUSSION

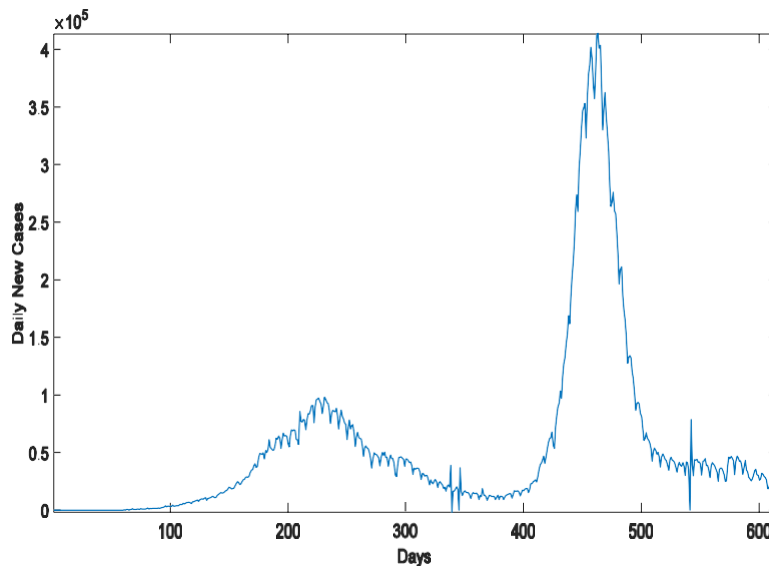


Figure 1: Plot of the time series Data

Recurrence plot the pan India COVID-19 daily new cases time series data is depicted in Figure 2.

The values of the Embedding Lag and Embedding Dimension for the data have been estimated and shown in Table 2.

Table 2: Embedding Lag and Embedding Dimension for the data

Embedding Lag	Embedding Dimension
33	3

The recurrence threshold  $\epsilon$  is a crucial parameter in the RP analysis. In this work recurrence threshold  $\epsilon$  is being chosen as 5% of the maximal phase space diameter [24].

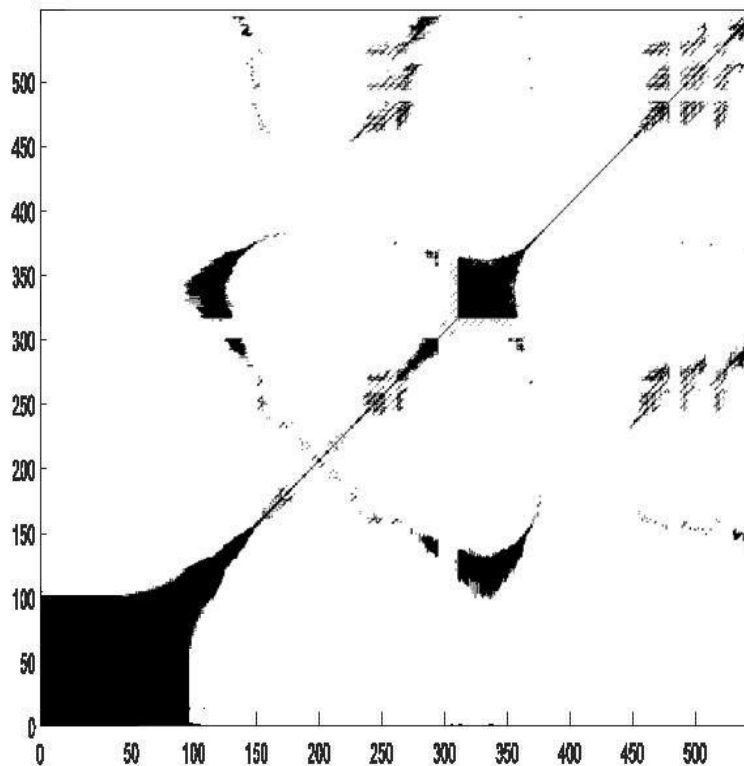


Figure 2: Recurrence plot the pan India COVID-19 daily new cases time series data

From the recurrence plot of data the following observations has been reported:

- a) In the first look the RP is quite similar with the RP of fractional Browning motion (fBm). This observation suggests that the signal behaves like a fractional Brownian motion and is non-stationary in nature. As a result a detailed scaling analysis in gross as well as in different scale window is required.
- b) The presence of white bands or disruptions in Figure 2 indicates non-stationarity in the process. This non-stationarity is due to the fact that the nature of the driving forces behind the pandemic outbreak fluctuations varies with time.
- c) Presence of very few periodic patterns like short or long diagonal lines demands very low cyclicities in the system process.
- d) The presence of very few diagonal lines parallel to the LOI (Line of Identity) declares that the evolution of states of the dynamics is similar at different times. This demands that the process is less deterministic in nature and hence forecasting is difficult.
- e) Presence of single isolated points indicates some heavy fluctuations in the process which may be due to sudden changes in dynamics like infection outbreak.
- f) The vertical, horizontal lines or clusters says that there may be few states change slowly with time or do not change at all; a signature of laminar states.

The values of the four parameters of Recurrence Quantification Analysis for the data have been estimated and shown in Table 3.

Table 3: The values of the four parameters of Recurrence Quantification Analysis

<b>Variables</b>	<b>Estimated values</b>
REC	0.1758
DET	0.9577
LMAX	192
ENTR	2.12

The high values of %DET= 95.77% claims that the pan India COVID-19 daily new cases time series data is deterministic in nature.

The LINEMAX value has been estimated as 192. This value can be considered as a moderate value when compared to the length of LOI. This value signifies that the data may have a moderate level of regularity and stability. This moderate value of LMAX may indicate a probable presence of both chaos and periodicity (may be a periodic chaos) in the system.

As shown in Table 3, the ENTR= 2.12. This value may be considered as a high value which implies that the fluctuation in the process is more complex in structure and hence more information is required to expose its complete dynamics.

## 5. CONCLUSION

In the present scope the pan India COVID-19 new cases in daily basis has been observed. The target is a time series data set with a span from 30.01.2020 to 04.10.2021. The following are the outcomes of the study.

- The daily spread of Covid-19 in India has characteristics of fractional Brownian motion.
- It is governed by non-stationary process.
- The signal has fair amount of irregularity in its profile.
- The signal is deterministic and stable having low periodicity.
- The dynamical structure is complex and requires more information to be revealed.
- Probable chaotic nature has been found in the signal.

The present realization of stable, deterministic, irregular and chaotic profile of daily new confirmed cases of COVID-19 in pan India level possibly indicates a variable growth rate (a temporal profile) of infection with alterable functional depictions at different slices of growth or decay and this supports the illustrations of Samadder and Ghosh [25]. The findings further suggest thorough analysis of the data or signal in connection with scaling, non-linearity, chaos, fractality and self-similarity.

## References

- [1] World Health Organization (2020). Coronavirus disease (COVID-19) Pandemic, WHO. [Online]. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> on 31st March 2020
- [2] mygov.in. [Online]. <https://www.mygov.in/covid-19>
- [3] R Gupta and S. K Pal. (2020) medRxiv. [Online]. <https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1>

- [4] R Gupta, G Pandey, P Chaudhary, and S. K Pal. medRxiv. [Online]. <https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1>
- [5] R Singh and R Adhikari. (2020) arXiv preprint arXiv:2003.12055. [Online]. <https://arxiv.org/pdf/2003.12055.pdf>
- [6] A Sahasranaman and N Kumar. (2020) SSRN. [Online]. <https://arxiv.org/ftp/arxiv/papers/2003/2003.09715.pdf>
- [7] J.H. Tanne et al., "Covid-19: how doctors and healthcare systems are tackling coronavirus world-wide.," *Bmj*, p. 368, 2020.
- [8] T Singhal, "A review of coronavirus disease-2019 (COVID-19).," *The Indian Journal of Pediatrics*, pp. 1-6, 2020.
- [9] C Sohrabi et al., "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19).," *International Journal of Surgery*, 2020.
- [10] Koushik Ghosh. (2020) SSRN. [Online]. <https://ssrn.com/abstract=3574066>
- [11] N Marwan and J Kurths, *Cross Recurrence Plots and Their Applications*, Eds.: C. V. Benton, Ed. Hauppauge: Nova Science Publishers, 2004.
- [12] J P Eckmann, S O Kamphorst, and D Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.*, vol. 5, p. 973, 1987.
- [13] N Marwan, M C Romano, and J Kurths, "Recurrence Plots for the Analysis of Complex Systems," *Phys. Rep.*, vol. 438, no. 5-6, pp. 237-329, 2007.
- [14] Rajdeep Ray, Mofazzal Hossain Khondekar, Koushik Ghosh, and Anup Kumar Bhattacharjee, "Complexity and periodicity of daily mean temperature and dew-point across India," *J. Earth Syst. Sci.*, vol. 128, no. 143, pp. 1-12, May 2019.
- [15] J P Eckmann, S O Kamphorst, and D Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.*, vol. 4, no. 9, pp. 973-977, 1987.
- [16] N Marwan, N Wessel, U Meyerfeldt, A Schirdewan, and J Kurths, "Recurrence plot based measures of complexity and its application to heart rate variability data," *Phys. Rev. E*, vol. 66, no. 2, pp. 1-8, 2002.
- [17] N Thomasson, T J Hoepfner, C L Jr. Webber, and J P Zbilut, "Recurrence quantification in epileptic EEGs," *Phys. Lett. A*, vol. 279, no. 1-2, pp. 94-101, 2001.
- [18] F Takens, "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, Rand David and Young Lai-Sang, Eds. Berlin: Springer, 1981, vol. 898, p. 366-387.
- [19] M S Roulston, "Estimating the errors on measured entropy and mutual information," *Physica D*, vol. 125, no. 3-4, p. 285-294, 1999.
- [20] M B Kennel, R Brown, and H Abarbane, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. Lett. A*, vol. 45, p. 3403-3411, 1992.
- [21] N Marwan, M C Romano, and J Kurths, "Recurrence Plots for the Analysis of Complex Systems," *Phys. Rep.*, vol. 438, p. 237, 2007.
- [22] J P Zbilut and C L Jr. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A*, vol. 171, no. 3-4, p. 199-203, 1992.
- [23] C L Jr. Webber and J P Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.*, vol. 76, no. 2, p. 965-973, 1994. [Online]. <http://www.ncbi.nlm.nih.gov/pubmed/8175612>
- [24] S Schinkel, O Dimigen, and N Marwan, "Selection of recurrence threshold for signal detection," *Eur. Phys. J. Special Topics*,

vol. 165, no. 1, p. 45–53, 2008.

- [25] Swetadri Samadder and Koushik Ghosh, "Analysis of Self-Similarity, Memory and Variation in Growth Rate of COVID-19 Cases in Some Major Impacted Countries," in *Journal of Physics: Conference Series*, 1797, 012010, p. Swetadri Samadder and Koushik Ghosh.



**Rajdeep Ray** received his B.Tech. degree in Electronics and Communication Engineering from REC, Durgapur, India, in 2002, and the M.Tech degree from National Institute of Technology, Durgapur, India, in 2008. He was awarded Ph.D in Electronics and Communication Engineering from National Institute of Technology, Durgapur, India, in 2020. Currently he is Assistant Professor in the Department of Electronics and Communication Engineering, Dr. B.C. Roy Engineering College, Durgapur. He has a total teaching experience of seventeen years. His research interest is in the area of Statistical Signal Processing, Nonlinear Dynamics and Time Series.



# DIFFERENTIAL EVOLUTION BASED MULTILEVEL HETEROGENEOUS CLUSTERING TECHNIQUE FOR IOT-BASED WIRELESS SENSOR NETWORKS

Sandip K Chaurasiya<sup>1\*</sup>, Arindam Biswas<sup>2</sup>, Rajib Banerjee<sup>3</sup>

<sup>1</sup>University of Petroleum & Energy Studies (UPES), Department of Cybernetics, School of Computer Science, Energy Acres Building, Bidholi, Dehradun, UK-248007, INDIA, Email: [schaurasiya@ddn.upes.ac.in](mailto:schaurasiya@ddn.upes.ac.in)

<sup>2</sup>School of Mines, Kazi Nazrul University, Asansol, West Bengal, INDIA, Email: [mailarindambiswas@yahoo.co.in](mailto:mailarindambiswas@yahoo.co.in)

<sup>3</sup>Department of Electronics & Communication Engineering, Dr. B. C. Roy Engineering College Durgapur, West Bengal, INDIA, Email: [rajib123banerjee@gmail.com](mailto:rajib123banerjee@gmail.com)

**Abstract** - The recent deployment of various intelligent and smart systems in Internet-of-Things (IoT) has been conceptualized from the wireless sensor network features for their actualization. The role of the deployed sensor nodes is quite significant for realizing the intended IoT applications for which sensor nodes are required to operate consistently. However, the very resource-constrained nature of sensor nodes imposes various limitations on the network performance. Out of these limitations, the one due to limited and irreplaceable energy resources is the most severe, which can down the network if not appropriately addressed. In this work, the issue mentioned above has been addressed via the formulation of energy-balanced clusters in an IoT-based heterogeneous wireless sensor network. The proposed clustering scheme takes various network parameters such as intra-cluster distances of the nodes and their respective energy status into account. This metaheuristic scheme formalizes the energy-balanced clusters independent of energy-heterogeneity deployed in the network. Moreover, the supremacy of the proposed scheme has been established in terms of network lifetime, network energy consumption, and network throughput over a state-of-the-art scheme via an extensive set of simulations and experimentation.

**Keywords** - *Clustering, Differential Evolution, Energy-Efficiency, Heterogeneity, Metaheuristic Techniques, Network Lifetime.*

## 1. INTRODUCTION

The Internet of Things (IoT) is the central technology for many of the cutting-edge applications such as intelligent healthcare applications, smart cities, environmental monitoring, and intelligent industrial applications. In each of the applications mentioned above, various devices sense and communicate their readings over the Internet to fulfill the intended objective(s). These IoT systems comprise the sensor nodes featured with limited power, memory, processing capabilities. In most applications, sensor nodes are left unattended after deploying IoT-enabled smart systems. Thus, the development of the schemes enabling energy-efficient network operations has emerged as a very critical problem among the researcher's community. A huge number of protocols have already been proposed operating at different layers of sensor networks like MAC, network, transport, and application layers. However, transmission among all the possible network operations consumes most of the nodes' energy. With this, the development of energy-efficient schemes emphasizing transmissions, particularly at the network layer, has attracted researchers. In this regard, clustering has emerged as the most widely used tool to enable energy-



efficient network layer operations and to lead to an enhanced network lifetime [1].

In a cluster-based network, nodes are primarily categorized as cluster-heads (CHs) and cluster-members. The network is divided into some clusters, and there could be exactly one cluster-head serving to the rest of the (member) nodes. In a cluster, member nodes are meant to sense the surroundings and generate the respective readings. Thereafter, they transfer their measurements to the respective cluster heads. The cluster head, in turn, collect the data from all its members, aggregate, and communicate the aggregated data to the base station (BS).

From the above discussion, it can be easily intuited that the cluster-heads perform mostly energy-intensive tasks compared to their member node counterparts. Since the cluster heads are selected from the same set of sensor nodes, the additional responsibilities being executed by the CHs might result in a quick drain of the battery leading to the early death of the nodes.

To address the problem described above, many works such as [2-4] have reported the introduction of specialized nodes, termed Gateways nodes. These nodes are provisioned with high energy with respect to the normal sensor nodes in the network and are designated as pre-configured cluster heads. However, the gateways nodes are also battery operated, and hence, they are required to be handled intelligently to perform for a longer duration. In the process of cluster formation, the deployed normal nodes are associated and administered by an appropriate gateway node based on metrics like received signal strength indicator (RSSI) and other scheme dependent parameters.

In the cluster formation process, the distribution of nodes among the gateway-led clusters matters a lot. It can be intuited that the poorly formed clusters (say overcrowded or sparse) may result in poor network performance, whereas load-balanced clusters might improve network performance. For example, for an overloaded cluster, the gateway node acting as cluster head would be overburdened in executing the operations like data receiving from its respective member nodes and transmitting them to the BS after performing data-aggregation in comparison to the peer gateway node with a sparse node distribution.

In addition to the deployment of specialized gateway nodes to facilitate the network operations, the network may also comprise heterogeneous nodes with different profiles and abilities [5]. This work considers a network in which the normal member nodes may be equipped with different initial power, termed energy-heterogeneity based heterogeneous wireless sensor network. More illustratively, the target application of this work is an IoT-based system with the supporting wireless sensor network consisting of energy heterogeneous nodes. The underlying sensor network consists of high-energy gateway nodes to act as the cluster heads, and the member nodes may contain any level of energy-heterogeneity. They all can be equipped with different powers, or all can have the same energy level depending on the very nature of application.

This work aims to achieve energy-efficient network operations like routing for the above-mentioned IoT-based wireless sensor networks via energy-balanced cluster formation. We have used a metaheuristic scheme, differential evolution, to meet this objective. Since the network may contain nodes with different energy levels, the proposed scheme Differential Evolution based Metaheuristic Heterogeneous Clustering Technique (DEMHCT) considers nodes' energy status and their proximity for the formulation of energy-balanced clusters independent of the deployed level of energy-heterogeneity. The DEMHCT proposes a suitable fitness function to partition the network in a finite number of clusters defined by the number of gateways deployed. An extensive set of simulations and experimentation demonstrate the establishment of the proposed scheme's supremacy over a state-of-the-art scheme in terms of improved network lifetime, throughput, and network energy consumption.

### **Major Contributions and Organization of Paper:**

The major contributions of the proposed DEMHCT are as follows:

- Devising a suitable fitness function that leads to
  - energy-balanced clusters formation.
  - reduced intra-cluster communication cost.
- Developing an energy-efficient clustering scheme based on the proposed fitness function.
- Performance analysis of the proposed scheme, DEMHCT under varying levels of energy-heterogeneity.

The remaining paper is organized in five successive sections such that section 2 narrates the related work. Section 3 briefs assumed network and energy-consumption models. Section 4 describes the proposed scheme in detail. Section 5 explains the simulation results with appropriate briefing, and finally, section 6 concludes the entire work.

## 2. RELATED WORKS

This section presents some of the significant contributions towards the development of energy-balanced clusters in the above-mentioned IoT-based wireless sensor network with gateways as follows:

In [4], the authors proposed a differential evolution-based clustering algorithm for the wireless sensor network for the above-discussed network architecture. The authors attempted the balanced clusters' formation via an appropriate fitness function in their proposed scheme. The proposed fitness function considers the gateway nodes' lifetime and the nodes' distance from their respective gateways. [4] uses binomial crossover and employs *DE/best/1* as its mutation strategy. A local improvement phase was also introduced, reassigned a node from the overloaded gateway to a gateway with comparatively low in each iterative step of the metaheuristic scheme. However, neither the fitness function nor the local improvement phase took the cluster size of cluster-length into account.

In a work [6], the authors proposed a genetic algorithm-based clustering solution for the aforesaid two-tier wireless sensor network (WSN) that is the network with traditional normal nodes and the energy-enriched specialized gateway nodes. In their proposed scheme, authors provisioned selection of the individuals on the basis of Roulette-Wheel selection strategy ensuring that the selection probability increases with the fitness value of the individuals. Here, the fitness function considers the network lifetime as the number of round until the first gateway node drains out of its battery. For the production of new offspring, k-point crossover or uniform crossover is implemented on a random basis. Moreover, to further improve an individual's fitness, mutation is applied where instead of random selection, a node is selected, dissipating the maximum energy due to transceiving its data. The scheme performs well over the traditional multihop schemes; however, provisioning only the network lifetime for the definition of its fitness function limits the scheme's performance.

In a work [7], the authors proposed a differential evolution-based memetic algorithm for achieving energy-efficient routing for the above-mentioned two-tier WSN. In their proposed fitness function, authors aimed to minimize gateway nodes' overall energy consumption in a

route starting at the gateway node and ending at the BS. [7] uses binomial crossover and employs *DE/best/2* as its mutation strategy. Moreover, to prove the scalability of the proposed scheme, they deployed more than a thousand energy-enriched gateway.

In one work [8], the authors proposed another clustering method named Gravitational Search Algorithm based Energy Efficient Clustering (GSA-EEC) for the WSNs. In the definition of the fitness function for their proposed scheme, GSA-EEC, authors considered the residual energy of the gateway nodes, the Euclidean distance between the gateway and sensor nodes, and the distance between sensor nodes and the base station. To prove the supremacy of the proposed scheme, the authors compared the scheme's performance to the algorithms like genetic algorithm and PSO-C in terms of network lifetime, throughput, and network energy consumption.

In [9], the authors proposed a differential evolution-based energy-efficient clustering method for the WSNs. The proposed scheme is basically aimed at the [4] for further improvement. In their proposed scheme, the authors updated the mutation function to accommodate the target vector along with the best vector and two other population vectors in comparison to the one utilized in [4]. More illustratively [9] uses binomial crossover and employs *DE/best/2(current-to-rand)* as its mutation strategy. However, the fitness function definition requires only the energy status of both the nodes- normal and the gateway.

In addition to the above discussed schemes, many works have also been proposed to address the energy-heterogeneity in the network such as [10–15]. However, all these schemes require the level of deployed energy-heterogeneity to be known a priori.

Considering the above works, it is very important to formalize energy-balanced clusters independent of the energy-heterogeneity deployed in the network. Moreover, additional importance to be given for maximizing resource utilization and reducing intra-cluster communication cost.

### 3. MODELS & ASSUMPTIONS

In this section, the various assumptions along with the network and energy-consumption models will be described.

#### A. Network Model

DEMHCT assumes the following network model:

- In the deployed IoT-based wireless sensor network, the underlying sensor network comprises two types of nodes- energy-enriched specialized gateway nodes and normal nodes.
- All the energy-intensive tasks are to be performed by the energy-enriched gateway nodes.
- The normal nodes may contain any level of energy-heterogeneity. For example, all the nodes may be equipped with the same initial energy or different initial energies.
- The normal nodes are distributed randomly across the sensing field and static in nature.

- The sensor nodes are equipped with the power control features.
- The base station (BS) is static and located at the center of the field.
- The sensor nodes periodically sense the environment in order to generate the data.

### B. Energy-Consumption Model

For the measurement of energy consumption by the nodes- gateway and normal nodes, the popular *First Order Radio Model* [16-19] is used here as follows:

$$E_{TX}(l, d) = \begin{cases} E_{elec} * l + \epsilon_{fs} * l * d^2, & d \leq d_0 \\ E_{elec} * l + \epsilon_{mp} * l * d^2, & d > d_0 \end{cases} \quad (1)$$

$$E_R(l) = E_{elec} * l \quad (2)$$

$$E_D(l) = \epsilon_{da} * l \quad (3)$$

where  $E_{TX}(l, d)$  refers to the energy required for the transmission of  $l$ -bits over a distance of  $d$  meters;  $E_{RX}(l)$  refers to the energy required for receiving  $l$ -bits message; and  $E_{DA}(l)$  denotes the energy required for aggregating  $l$ -bits message.  $\epsilon_{fs}$  and  $\epsilon_{mp}$  are the energy required to run the amplification circuitry in free-space and multipath fading models respectively.  $\epsilon_{da}$  is the energy required to aggregate a single bit information.  $d_0$  is the threshold distance adapted by the first order radio model based on  $\epsilon_{fs}$  and  $\epsilon_{mp}$  as follows:

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \quad (4)$$

### C. Mutation & Crossover Strategy

The proposed scheme, DEMHCT, uses *DE/best/1* as its mutation strategy and calls *binomial crossover* for the intended crossover purpose.

## 4. PROPOSED SCHEME- DIFFERENTIAL EVOLUTION BASED MULTILEVEL HETEROGENEOUS CLUSTERING TECHNIQUE (DEMHCCT)

This section explains the proposed scheme, Differential Evolution based Multilevel Heterogeneous Clustering Technique (DEMHCCT) for the network depicted in subsection 3-A. Like its predecessors, DEMHCCT is also a base station assisted clustering scheme which calls the BS to decide the suitable partitioning of the network into a finite number of clusters induced by the gateways deployed. After the cluster formation, BS transfers the clusters' control to the respective gateways for further network operations.

The proposed scheme comprises two phases- clustering and steady-state phase.

**Clustering Phase:** In the clustering phase, the sensor nodes broadcast their unique IDs to be sensed by the gateways nodes in their communication range. The gateway nodes communicate such collected IDs to the centralized base station to form energy-balanced clusters. Once the clusters are formed, the gateways are notified of their respective cluster members. Then, the gateway nodes prepare and communicate the TDMA schedule to their respective member nodes.

Furthermore, the data transmission may then take place between the nodes & gateways, and gateways and BS accordingly in the steady-state phase.

As stated in subsection 3-C, DEMHCT employs *DE/best/1/bin* for the formation of energy-balanced clusters in the network. Like in any metaheuristic scheme, the clustering process starts with the initialization phase, wherein a finite set of population vectors is created by assigning the complete set of network nodes to the available gateways. Afterward, the mutation and crossover phases are called iteratively for a definite number of time to refine the population vectors accordingly as per the devised fitness function. At the end of the predefined number of iteration, the vector with the best fitness value is chosen as the final clusters' specification [20].

For the success of any metaheuristic scheme, the design of an appropriate fitness function is always a key. Keeping the same in view, DEMHCT derives a fitness function ( $\Phi$ ) that considers multiple parameters such as residual energy of the nodes & gateways, energy distribution among the tentative clusters, and intra-cluster distance of the nodes from the respective gateway.

In the derivation of the fitness function, it has been assumed that a higher fitness value indicates,

- minimum standard deviation of gateways' lifetime.
- minimum standard deviation of clusters' average distance.
- minimum standard deviation of cluster's energy.

On the basis of the above ideation, the fitness function ( $\Phi$ ) can be given as follows:

$$\Phi = \frac{k\sqrt{k}}{\sqrt{(\mu_{L_{GW}} - L_{GW_i})^2 * (\mu_D - AvgCDist_{GW_i})^2 * (IdealE - E_{CL_i})^2}} \quad (5)$$

where,  $k$  is the number of gateways deployed;  $L_{GW_i}$  refers to the expected lifetime of  $i^{th}$  gateway defined as ratio of gateway's residual energy to the tentative energy consumption by the gateway node in carrying out the assigned duty of cluster data aggregation and transmission to BS;  $\mu_{L_{GW}}$  is the mean gateway lifetime;  $AvgCDist_{GW_i}$  is the average distance of the member nodes from their respective gateway in  $i^{th}$  cluster;  $\mu_D$  is the average of average cluster distances for all the clusters;  $IdealE$  is the ideal cluster energy which is anticipated in a cluster towards the even distribution network energy among the clusters; and,  $E_{CL_i}$  is the current energy of the  $i^{th}$  cluster. More illustratively,

$$L_{GW_i} = \frac{ResidualEnergy_{GW_i}}{ClusterSize_i * (E_{RX} + E_{DA}) + E_{TX}(GW_i, BS)} \quad (6)$$

$$\mu_{L_{GW}} = \frac{1}{k} * \sum_{i=1}^k L_{GW_i} \quad (7)$$

$$AvgCDist_{GW_i} = \frac{1}{m} * \sum_{j=1}^m distance(GW_i, node_j) \quad (8)$$

where,  $m$  is the number of member nodes in the  $i^{th}$  cluster.

$$\mu_D = \frac{1}{k} * \sum_{i=1}^k AvgCDist_{GW_i} \quad (9)$$

$$IdealE = \frac{1}{k} * \sum_{i=1}^N ResidualEnergy_{Node_i} \quad (10)$$



To prove the supremacy of the DEMHCT over a state-of-the-art scheme [9], an extensive set of experiments has been conducted. In executing the experimentation, the underlying wireless sensor network with 100 normal nodes and 15 gateways for the intended IoT application has been deployed in a field of  $100 \times 100 \text{ m}^2$ . The BS is positioned at the center of the sensing field (50m, 50m). The normal nodes might be configured with arbitrary levels of energy-heterogeneity as per the requirement of the deployed application. A set of exemplary network deployment for 3-level and 5-level energy-heterogeneities has been presented in Figure 1. The nodes labeled with 'G' refer to the gateways, and that labeled with 'BS' refers to the base station. Every other node is representing the normal nodes in the two-tier IoT-based wireless sensor network.

As stated earlier in subsection 3-B, the proposed scheme, DEMHCT adopts First Order Radio Model to compute energy consumption in a variety of network operations. The simulations have been conducted here in MATLAB.

Moreover, a complete set of parameters' description is given in Table 1 as follows:

Table 1 - Network Parameters Used in the Simulation

Parameter	Parameter's Value
Network Area	$100 \times 100 \text{ m}^2$
Base Station's Position	(50m, 50m)
Number of Nodes Deployed in the Network	100
Energy-Heterogeneity Deployed in the Network	{1L, 2L, 3L, 4L, 5L}
Initial Energy of the Normal Nodes	2J
Energy Multipliers for {1L, 2L, 3L, 4L, 5L}	{0, 2.0, 3.5, 4.5, 5.5}
Proportional Factors	{0.8, 0.6, 0.7, 0.8}
Size of Data Message	4000 bits
Energy Consumption in Data Aggregation ( $\epsilon_{da}$ )	5 nJ/bits/signal
Energy Consumption in Running the Transceiver's Circuitry ( $E_{elec}$ )	50 nJ/bit
Amplification Factor- Free Space Model ( $\epsilon_{fs}$ )	10 pJ/bit/m <sup>2</sup>
Amplification Factor- Multipath Fading Model ( $\epsilon_{mp}$ )	0.0013 pJ/bit/m <sup>4</sup>
Mutation Factor (F)	1.75
Crossover Factor	0.8

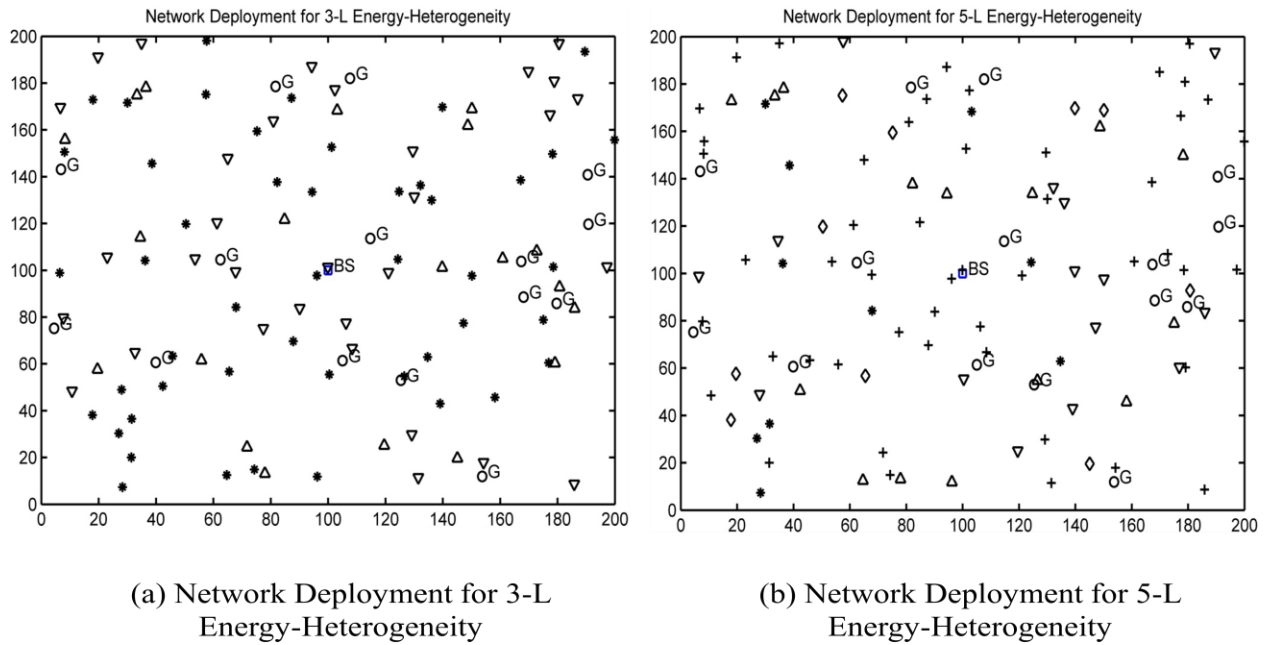


Figure 1 - Simulation Interface For Network Operation

## B. Simulation & Experimentation

This subsection describes the various simulation results obtained through extensive experiments. Experiments are performed to establish the supremacy of the proposed scheme, DEMHCT over [9] in terms of:

- Improved network lifetime, decreased nodes' death rate, reduced network energy consumption, and enhanced throughput at the base station.
- Sustainability of DEMHCT under any level of energy-heterogeneity.

Figure 2 depicts the outperformance of DEMHCT over [9] in terms of improved network lifetime. Here, lifetime is defined as the time when the first gateway dies in the network [4, 6, 7, 9]. Figure 2 clearly state that irrespective of the deployed level of energy-heterogeneity, DEMHCT always performs better than [9]. The scheme of [9] has been referred to here by the name "Milad" for notational convenience.

The tabular comparison exhibited in Table 2 can be consulted for more clarity of the aforesaid. It can be easily intuited from the Table 2 that DEMHCT claims 39.94%, 20.91%, 12.86%, 9.5%, & 14.64% gain in terms of network lifetime over [9] for 1-L, 2-L, 3-L, 4-L, & 5-L of energy-heterogeneity respectively.



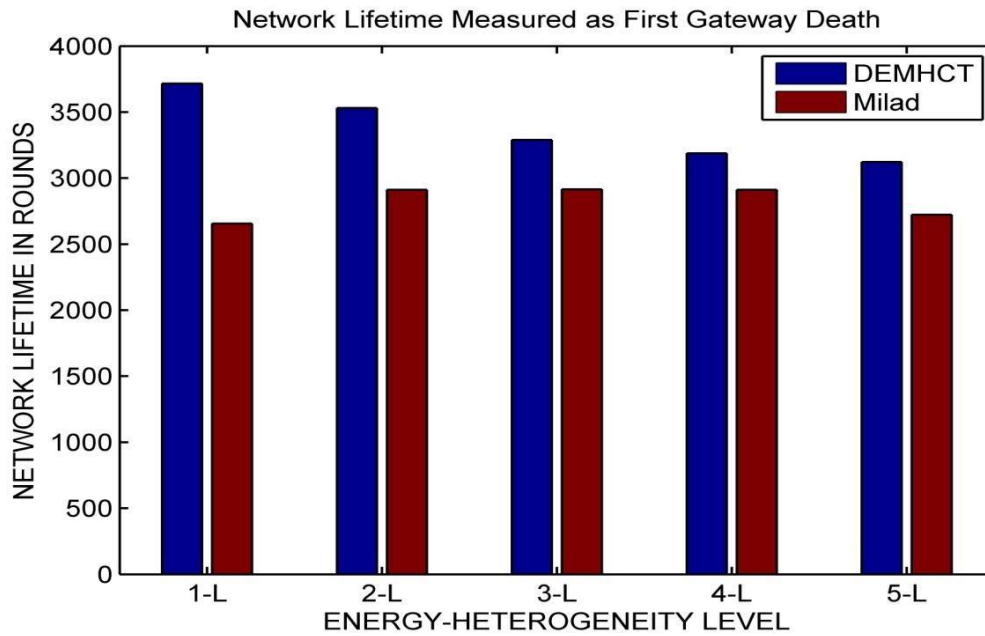
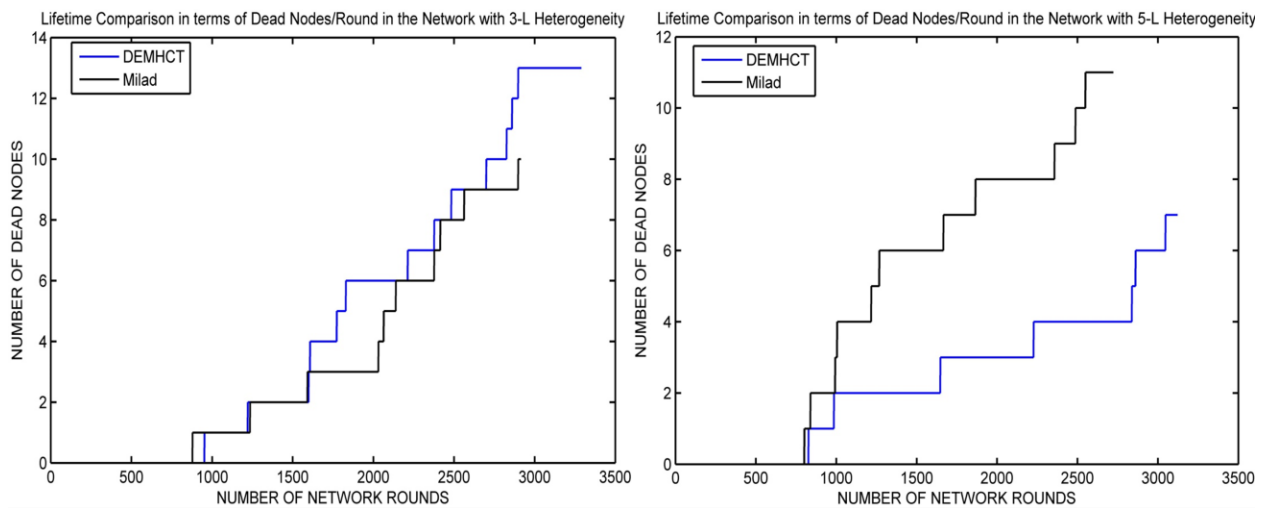


Figure 2- First Gateway Death in the Network

Table 2- First Gateway Death Events

Scheme\Heterogeneity	Level-1	Level-2	Level-3	Level-4	Level-5
DEMhCT	3717	3531	3290	3189	3123
Milad	2656	2912	2915	2912	2724

Figure 3(a) & 3(b) confirm the outplay of DEMHCT with respect to [9] in terms of nodes' death rate over the network rounds in the presence of 3-level and 5-level energy-heterogeneity, respectively.



(a) Nodes' Decay Rate for 3-L Energy-Heterogeneity

(b) Nodes' Decay Rate for 5-L Energy-Heterogeneity

Figure 3- Nodes' Death Rate over the Network Rounds

Table 3 and Table 4 describe the comparative performance of both the schemes DEMHCT and [9] in terms of the average residual energy/node and network energy consumption, respectively. The outperformance of DEMHCT over [9] is evident from both the aforesaid tables.

Table 3 depicts that for the 1-level (homogeneous), 2-level, 3-level, 4-level, and 5-level energy-heterogeneities among the normal nodes, DEMHCT enables the nodes with 13.65%, 9.92%, 9.92%, 7.03%, & 13.34% more average residual energy over the respective network lifetime with respect to [9].

Similarly, Table 4 describe that DEMHCT successfully utilizes more network resource available as network energy in comparison to [9] for each level of energy-heterogeneity.

Table 3- Average Residual Energy per Node over the Network Lifetime

Heterogeneity	$Lifetime = \max(FGD_{DEM\text{HCT}}, FGD_{Milad})$		$Lifetime = \min(FGD_{DEM\text{HCT}}, FGD_{Milad})$	
	DEM HCT	Milad	DEM HCT	Milad
1-L	0.8588	0.7556	1.2020	1.0575
2-L	3.7982	3.4552	4.6059	4.1899
3-L	5.2213	4.7805	5.8930	5.3955
4-L	7.2320	6.7568	7.9201	7.3997
5-L	8.4723	7.4755	9.7138	8.5708

Table 4- Network Energy Consumption as Resource Utilization

Scheme\Heterogeneity	Level-1	Level-2	Level-3	Level-4	Level-5
DEM HCT	281.0586	352.2352	367.5005	336.2212	324.3492
Milad	229.8225	276.7929	328.1614	299.1633	289.3648

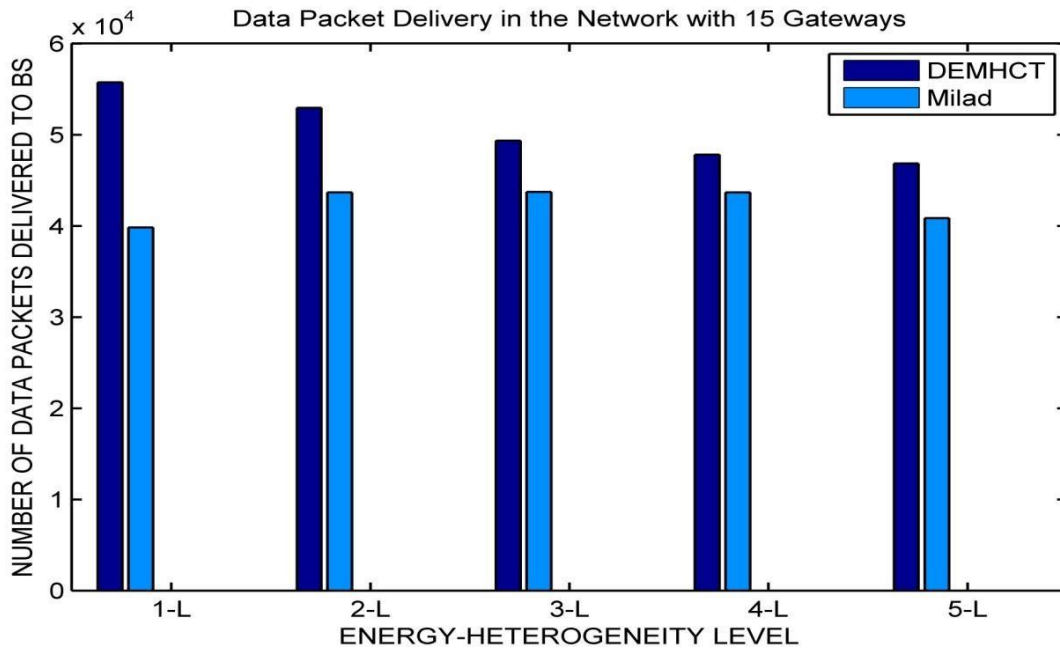


Figure 4- Data Packet Delivery at Base Station

Figure 4 depicts the supremacy of the proposed scheme, DEMHCT over [9] in terms of the number of data packets delivered to the BS under varying levels of energy heterogeneity. It also

indicates that irrespective of the deployed level of energy-heterogeneity, DEMHCT always performs better than [9].

Table 5- Data Packet Delivery at Base Station

Scheme\Heterogeneity	Level-1	Level-2	Level-3	Level-4	Level-5
DEMhCT	55740	52950	49350	47820	46830
Milad	39825	43665	43725	43665	40845

Moreover, Table 5 describes the same quantitatively. It can be easily intuited from the Table 5 that DEMHCT claims 39.96%, 21.26%, 12.86%, 9.5%, & 14.65% gain in terms of network throughput over [9] for 1-L, 2-L, 3-L, 4-L, & 5-L of energy-heterogeneity respectively.

Thus, the better performance of the proposed scheme, DEMHCT over [9] is evident from the above-obtained simulation results in figures 2 - 4 and tables 2 - 5.

## 6. CONCLUSION & FUTURE WORK

In this work, a differential evolution based metaheuristic clustering technique (DEMhCT) has been proposed for a two-tier IoT-based wireless sensor network. The two-tier network considered here employs two different categories of nodes- energy-enriched gateway nodes and environment sensing normal nodes. In addition to the architecture mentioned above, the network may contain any level of energy-heterogeneity for its normal nodes. To achieve energy-efficient network operations, clustering is employed. The proposed scheme formulates a specialized fitness function to ensure network partitioning into a finite number of energy-balanced clusters by considering gateways' and nodes' energy, clusters' energy and intra-cluster distance appropriately. Through an extensive set of simulations and experimentation, the performance of the proposed scheme, DEMhCT, has been established in terms of improved network lifetime, enhanced average residual energy, and increased data packet delivery.

An IoT-based WSN with mobile nodes would be investigated in future as an extension of this work.

## References

- [1] Gupta, G.P., Saha, B. Load balanced clustering scheme using hybrid metaheuristic technique for mobile sink based wireless sensor networks. *J Ambient Intell Human Comput* (2020). <https://doi.org/10.1007/s12652-020-01909-z>.
- [2] G. Gupta, M. Younis, Load-balanced clustering of wireless sensor networks, in: *IEEE International Conference on Communications, ICC'03*, vol. 3, IEEE, 2003, pp.1848–1852.
- [3] C.P. Low, C. Fang, J.M. Ng, Y.H. Ang, Efficient load-balanced clustering algorithms for wireless sensor networks, *Comput. Commun.* 31 (4) (2008) 750–759.
- [4] Pratayay Kuila, Prasanta K. Jana, A novel differential evolution based clustering algorithm for wireless sensor networks, *Applied Soft Computing*, 25, 2014, pp. 414-425.

- [5] Dumka, A., Chaurasiya, S. K., Biswas, A., and Mandoria, H. L, A Complete Guide to Wireless Sensor Networks: from Inception to Current Trends, 1st Edition; CRC Press, Boca Raton, Florida, USA, 2019.
- [6] A. Bari, S. Wazed, A. Jaekel, S. Bandyopadhyay, A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks, *Ad Hoc Netw.* 7 (4), 2009, pp. 665–676.
- [7] U.K. Chakraborty, S.K. Das, T.E. Abbott, Energy-efficient routing in hierarchical wireless sensor networks using differential-evolution-based memetic algorithm, in: 2012 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2012, pp. 1–8.
- [8] Rao, PC Srinivasa, Haider Banka, and Prasanta K. Jana. "Energy efficient clustering for wireless sensor networks: A gravitational search algorithm." In International Conference on Swarm, Evolutionary, and Memetic Computing, pp. 247-259. Springer, Cham, 2015.
- [9] M. Ghahramani and A. Laakdashti, Efficient Energy Consumption in Wireless Sensor Networks Using an Improved Differential Evolution Algorithm, 10th International Conference on Computer and Knowledge Engineering (ICCKE), 2020, pp. 018-023, doi: 10.1109/ICCKE50421.2020.9303713.
- [10] L. Qing, Q. Zhu, M. Wang, Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *ELSEVIER, Computer Communications* 29, 2006, pp 2230- 2237, DOI: <https://doi.org/10.1016/j.comcom.2006.02.017>.
- [11] Elbhiri, B. , Saadane, R. , El Fkihi, S. , Aboutajdine, D., Developed Distributed Energy-Efficient Clustering (DDEEC) for heterogeneous wireless sensor networks, in: 5th International Symposium on I/V Communications and Mobile Network (ISVC), 2010, DOI: 10.1109/ISVC.2010.5656252.
- [12] Parul Saini, Ajay. K. Sharma, E-DEEC- Enhanced Distributed Energy Efficient Clustering Scheme for heterogeneous WSN, in: 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010), pp. 205-210, DOI: 10.1109/PDGC.2010.5679898.
- [13] N. Javaid, T. N. Qureshi, A.H. Khan, A. Iqbal, E. Akhtar, M. Ishfaq, EDDEEC: Enhanced Developed Distributed Energy-Efficient Clustering for Heterogeneous Wireless Sensor Networks, *Procedia Computer Science* 19 ( 2013 ) 914 – 919, DOI: <https://doi.org/10.1016/j.procs.2013.06.125>.
- [14] S. Chand, S. Singh, B. Kumar, Heterogeneous HEED protocol for wireless sensor networks, *Wireless Pers. Commun.* 77 (3) (2014) 2117–2139, DOI: <https://doi.org/10.1007/s11277-014-1629-y>.
- [15] S. Singh, Aruna Malik, Rajeev Kumar, Energy efficient heterogeneous DEEC protocol for enhancing lifetime in WSNs, *Engineering Science and Technology, an International Journal* 20 (2017) 345–353, DOI: <https://doi.org/10.1016/j.jestch.2016.08.009>.
- [16] Sandip Kumar Chaurasiya, Tumpa Pal, Sipra Das Bit, An Enhanced Energy-Efficient Protocol with Static Clustering for WSN, *Proceedings IEEE Xplore, Int'l Conf. of Information Networking (ICOIN)*, Kuala Lumpur, Malaysia, March 2011, pp. 58-63, DOI: 10.1109/ICOIN.2011.5723134.
- [17] Sandip K Chaurasiya, Jaydeep Sen, Srirupa Chatterjee, and Sipra Das Bit, EBLEC: An Energy-Balanced Lifetime Enhancing Clustering for WSN, *Proceeding IEEE Xplore, 14th Int'l Conf. on Advanced Communication & Technology-2012*, PyeongChang, Korea (South), Feb. 2012, pp. 189-194, INSPEC Accession Number: 12656578.
- [18] Sandip K Chaurasiya, Joydeep Mondal, Suman Datta, Field-of-View based hierarchical clustering to prolong network lifetime of WMSN with obstacles, *Proceeding IEEE Xplore, Int'l Conf. on Electronics, Communication, and Computational Engineering*, Chennai, INDIA, Nov. 2014, pp.72-77, DOI: 10.1109/ICECCE.2014.7086638.
- [19] Rajib Banerjee, Sibashis Chatterjee, Sipra Das Bit, An energy saving audio compression scheme for wireless multimedia sensor networks using spatio-temporal partial discrete wavelet

transform, *Computers and Electrical Engineering* 48, 2015, pp. 389–404, DOI: <http://dx.doi.org/10.1016/j.compeleceng.2015.09.009>.

[20] Sandip K Chaurasiya, Arindam Biswas, & Rajib Banerjee, "Metaheuristic Load- Balancing Based Clustering Technique for Wireless Sensor Networks", *Hindawi, Wireless Communication and Mobile Computing*, vol. 2022, Article ID 8911651, 21 pages, DOI: <https://doi.org/10.1155/2022/8911651>



## DIFFERENTIATING CODING FROM NON-CODING OF DNA SEQUENCES – A STATISTICAL APPROACH

Shashi Bajaj Mukherjee<sup>1\*</sup>, Pradip Kumar Sen<sup>2</sup>

<sup>1</sup>Department of B.Sc. & Hu.(Mathematics), Dr. B C Roy Engineering College, Durgapur- 713 206

Email: [sashibajaj.mukherjee@bcrec.ac.in](mailto:sashibajaj.mukherjee@bcrec.ac.in)

<sup>2</sup>Department of Mathematics, Jadavpur University, Kolkata-700 032

**Abstract-** Gene finding techniques are based on identifying coding portions in DNA sequences. In this paper, the difference between coding and non-coding sequences is studied and a criterion is developed to classify a sequence into one of the two types. The species considered in *Drosophila Melanogaster*, and all six chromosomes are studied individually. The basic characteristic of a sequence that is considered is the distribution of the four bases A, C, G and T. A statistical test of homogeneity is used to study the difference among the types in this respect. In this connection, we also tested the GC-richness property of coding sequences, which is validated. Finally, a distance type parameter is developed for each sequence. Actually  $d^2$ , the square of the Euclidean distance of the distribution of bases from a standard distribution is considered. For the standard distribution, various choices are examined. The frequency distribution of  $d^2$  for the coding and non-coding types shows clear differences for all choices for the standard distribution. Finally a critical value of  $d^2$  is obtained to classify a sequence as coding or non-coding, which gives more than 90% accuracy for all six chromosomes.

**Keywords:** DNA, Coding and non-coding sequences, compositional bias, discriminating statistic

\*Corresponding Author: Shashi Bajaj Mukherjee, Email: [sashibajaj.mukherjee@bcrec.ac.in](mailto:sashibajaj.mukherjee@bcrec.ac.in)

### 1. INTRODUCTION

In the evening of Saturday, 28th February, 1953 Crick announced in the presence of Watson that he had “discovered the secret of life”. He referred to the double helical structure of DNA [3]. Since then the study of DNA structure became on the focal point in the field of Molecular Biology. Significant advances have been made in areas such as DNA transcription, replication, protein synthesis etc., and many details of their inner workings have been revealed. However, there remain open questions in areas like working methodology of splicing, origin of INTRONS and EXONS, alternative splicing etc.

The purpose of this paper is to develop methods related to gene identification based on differences between INTRONS and EXONS In a DNA sequence only the coding portion i.e., EXONS can take part in carrying a trait. But 90% of whole sequences are INTRONS, which were previously considered as garbage. Recent studies indicate otherwise, though the exact function of INTRONS is unknown. A significant amount of research work has been done to differentiate the coding and non-coding parts of DNA sequences. Some of the standard techniques are Hidden Markov Models, Neural Networks (GRAIL), Grammars, Homology etc ([5], [6]).

Here the aim is to study the difference between coding and non-coding sequences of a genome

using statistical analysis. This type of approach, though it makes obvious sense, is relatively uncommon. The basic idea of Compositional bias, i.e., the differences in the distribution of the bases a, c, g, t, however, has been studied before ([1], [2], [4]).

## 2. DATA

For this analysis we have used data from GenBank where a large database of genomic sequences of various species has been stored. The considered species is *Drosophila Melanogaster* (Fruit Fly). The species has been chosen as its genomic sequence is considered reliable and is commonly investigated by researchers.

There are four chromosomes (CHR) of this genome, e.g. CHR2, CHR3, CHR4 and CHRX. The first two are further divided into two parts 2L, 2R and 3L, 3R respectively. GenBank provides the data in various formats, but here taken the “.gbk” format only which is the most suitable for this work.

## 3. PRELIMINARY PROCESSING

First have to divide the genomic sequences of each chromosome into five parts, to be stored in five files, which are similar to original “.gbk” genomic sequences file, but without the ORIGIN portion for the whole base sequence. Two files contain coding sequence, further classified into two types:

- i) CDS A (the continuous coding sequences),
- ii) CDS B (the coding sequences interrupted by INTRON A's).

The next two files contain non-coding sequences. They are also of two types:

- i) INTRON A (the non-coding sequences embedded in CDS B's),
- ii) INTRON B (these is also INTRON but not embedded in the CDS B's).

Though main goal is to differentiate coding and non-coding sequences, it would be interesting to study the difference, if any, between the two coding types and between the two non-coding types. The fifth file is formed by the ORIGIN part of the original file.

With each of the first four files using the fifth create fasta format files. These fasta files are basis for the next part of the analysis. The first step of analysis uses the relative frequencies of the four nucleotides in the four types of sequences. In the next step, from these fasta files created four respective base count files by counting the a's, c's, g's and t's in each of the individual sequences.

## 4. TEST OF HOMOGENEITY

Before embarking on main work, one question which comes to our mind is whether the four types of sequences differ with respect to the distribution of four nucleotides or not. This can be decided on the basis of the standard test of homogeneity [7]. The following table gives the data format for the statistical test. Here the  $f_{ij}$ 's are the total frequencies for a, c, g, t respectively for all the sequences combined for the four types and  $n_i, n_j$  are the respective column and row totals.

**Table 1: Test of Homogeneity**

Type Base	A	C	G	T	TOTAL
CDS A	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$n_1$
CDS B	$f_{21}$	$f_{22}$	$f_{23}$	$f_{24}$	$n_2$
INTRON A	$f_{31}$	$f_{32}$	$f_{33}$	$f_{34}$	$n_3$
INTRON B	$f_{41}$	$f_{42}$	$f_{43}$	$f_{44}$	$n_4$
TOTAL	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n$

The test Statistic is  $D = n \left[ \sum_{i=1}^4 \sum_{j=1}^4 \frac{f_{ij}^2}{n_i n_{.j}} - 1 \right]$

Which has a  $\chi^2$ -distribution with  $(4-1)(4-1) = 9$  degrees of freedom.

Critical values for  $\chi^2$ -distribution with 9 degrees of freedom are: 21.666 (for 1% level) and 16.919 (for 5% level).

The analysis may carry further by comparing coding vs. non-coding, CDS A vs. CDS B, and INTRON A vs. INTRON B also. The  $\chi^2$ - statistic in these three cases has 3 degrees of freedom and the corresponding critical values are 11.345 (1% level) and 7.815 (1% level)

The results found for Chromosome 2L are given in Table2.



**Table 2: Test of Homogeneity (CHR 2L)**

TEST OF HOMOGENEITY				
165661	162095	164476	130951	623183
1411445	1445760	1443601	1161613	5462419
3382254	2277948	2193517	3489488	11343207
5526987	3820545	3830613	5493678	18671823
10486347	7706348	7632207	10275730	36100632
The chi-squared value = 353671.655548				
Significant at 1% level.				
1577106	1607855	1608077	1292564	6085602
8909241	6098493	6024130	8983166	30015030
10486347	7706348	7632207	10275730	36100632
The chi-squared value for coding vs. non-coding = 216488156.694866				
Significant at 1% level.				
165661	162095	164476	130951	623183
1411445	1445760	1443601	1161613	5462419
1577106	1607855	1608077	1292564	6085602
The chi-squared value for coding sequence = 180.585748				
Significant at 1% level.				
3382254	2277948	2193517	3489488	11343207
5526987	3820545	3830613	5493678	18671823
8909241	6098493	6024130	8983166	30015030
The chi-squared value for non-coding sequence = 9725.510991				
Significant at 1% level.				

All the tests for each chromosome are significant at level 1% except in chromosomes 2R and X for CDS A vs. CDS B which are significant at level 5% or more. It may conclude that the four types of sequences are not at all homogeneous with respect to the distribution of nucleotides. Even the coding parts are not homogeneous, nor are the non-coding parts.

## 5. TEST OF G-C RICHNESS

G-C richness is a very commonly studied property of coding sequences. Essentially it means that coding sequences have higher combined frequencies of G and C bases than the non-coding sequences. For the test of G-C richness the following table has been considered as data table for the using test statistics.

**Table 3: Test of G-C richness**

Sl. No.	Type	A+T	C+G	TOTAL
1	CDS A	$n_1-x_1$	$x_1$	$n_1$
2	CDS B	$n_2-x_2$	$x_2$	$n_2$
3	INTRON A	$n_3-x_3$	$x_3$	$n_3$
4	INTRON B	$n_4-x_4$	$x_4$	$n_4$

These are tests of comparison of two proportions. In each case two proportions  $P$  are used to calculate the test statistic  $z$ , which has approximately a standard normal distribution for large sample size.

(I) Test of G-C richness for coding (i.e., CDS A+CDS B) vs. non-coding (i.e., INTRON A +INTRON B):

$$P_c = \frac{x_1 + x_2}{n_1 + n_2}, \quad P_n = \frac{x_3 + x_4}{n_3 + n_4}$$

$$z = \frac{P_c - P_n}{\sqrt{P(1-P)} \sqrt{\frac{1}{n_1 + n_2} + \frac{1}{n_3 + n_4}}}, \text{ where } P = \frac{x_1 + x_2 + x_3 + x_4}{n_1 + n_2 + n_3 + n_4}$$

This one sided test is significant at 1% level if  $z > 2.326$  and at 5% level if  $z > 1.645$ .

(II) Test of GC-richness for CDS A vs. CDS B:

$$P_1 = \frac{x_1}{n_1}, \quad P_2 = \frac{x_2}{n_2}$$

$$z = \frac{|P_1 - P_2|}{\sqrt{P(1-P)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ where } P = \frac{x_1 + x_2}{n_1 + n_2}$$

(III) Test of GC-richness for INTRON A vs. INTRON B:

$$P_3 = \frac{x_3}{n_3}, \quad P_4 = \frac{x_4}{n_4}$$

$$z = \frac{|P_3 - P_4|}{\sqrt{P(1-P)} \sqrt{\frac{1}{n_3} + \frac{1}{n_4}}}, \text{ where } P = \frac{x_3 + x_4}{n_3 + n_4}$$

In (II) and (III), the test is significant at 1% level if  $|z| > 2.58$  and at 5% level if  $|z| > 1.96$ .

In almost all cases (three cases for each of five chromosomes) the statistic is significant at 1% level. The exceptions are exactly those for the Homogeneity test i.e., CDS A vs. CDS B for chromosomes 2R and X. For example, for CHR 2L, the analysis is follows as in Table4:

**Table 4: Test of G-C richness (CHR 2L)**

TEST OF G-C RICHNESS		
2869670	3215932	6085602
17892407	12122623	30015030
The GC-Richness for coding vs. non-coding sequence = 566.818695		
Significant at 1% level.		
296612	326571	623183
2573058	2889361	5462419
The GC-Richness for coding sequence = 7.364723		
Significant at 1% level.		
6871742	4471465	11343207
11020665	7651158	18671823
The GC-Richness for non-coding sequence = 84.306506		
Significant at 1% level.		

The divided-bar graphs in Fig1 facilitate a clear visualization of the distribution of nucleotides for the four types of sequences for each chromosome and the G-C richness property:

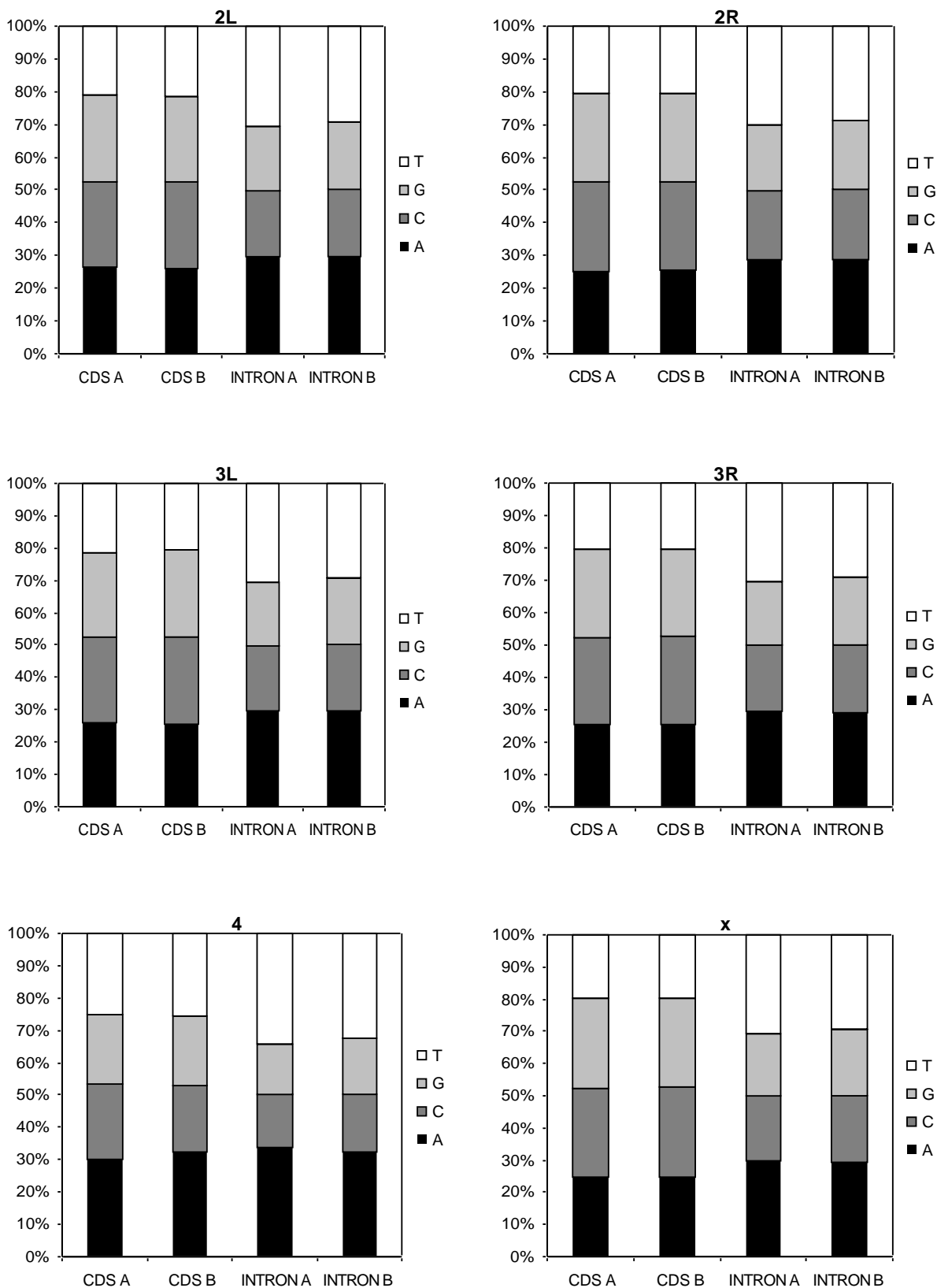


Fig 1: Distribution of the four nucleotides for the four types of sequences

## 6. DIFFERENTIATING CRITERION

For a more detailed analysis of the difference among the four types of sequences we use a distribution of a statistic similar to a chi-square statistic for goodness of fit tests. The standard chi-square statistic is

$$\sum_i \frac{(\text{ObservedVdue} - \text{ExpectedVdue})^2}{\text{ExpectedValue}}$$

$$= \sum_i \frac{(nP_i - n\pi_i)^2}{n\pi_i}$$

Where,  $P_i$  = Observed Proportion,  $\pi_i$  = Theoretical Proportion,  $n$  = Sample Size.

To avoid the huge scaling effect of the large sequence size  $n$  following statistics applied.

$$d^2 = \sum_{i=1}^4 (P_i - \pi_i)^2$$

We are going to use it as a discriminating statistic. We can consider this statistic as the squared Euclidian distance between the distributions  $\{P_i\}$  and  $\{\pi_i\}$ .

We use five different choices for the choice of  $\{\pi_i ; i = 1,2,3,4\}$ :

C<sub>1</sub>: Overall distributions for CDS A.

C<sub>2</sub>: Overall distributions for CDS B.

C<sub>3</sub>: Overall distributions for INTRON A.

C<sub>4</sub>: Overall distributions for INTRON B.

C<sub>5</sub>: Overall distributions for all four types combined.

Now we are going to construct the frequency distribution of  $d^2$  for each of the five cases.

Figures 2-7 represent the comparison of the four types of sequences in terms of frequency distributions of  $d^2$  for the six chromosomes. In each case the horizontal axis represents five hundred times  $d^2$  and the five charts relate to the cases C<sub>1</sub> to C<sub>5</sub> respectively.

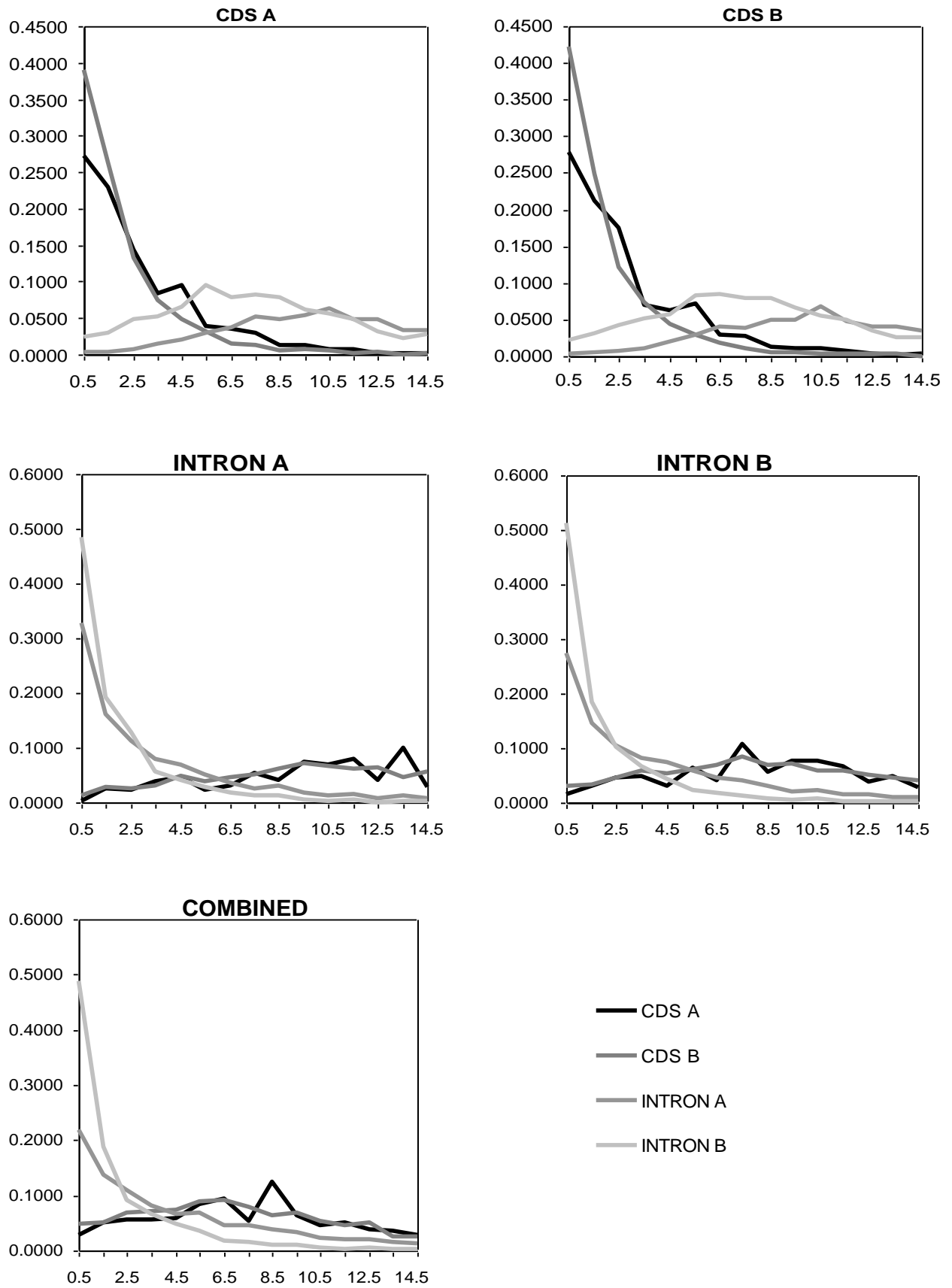


Fig 2: Chromosome 2L

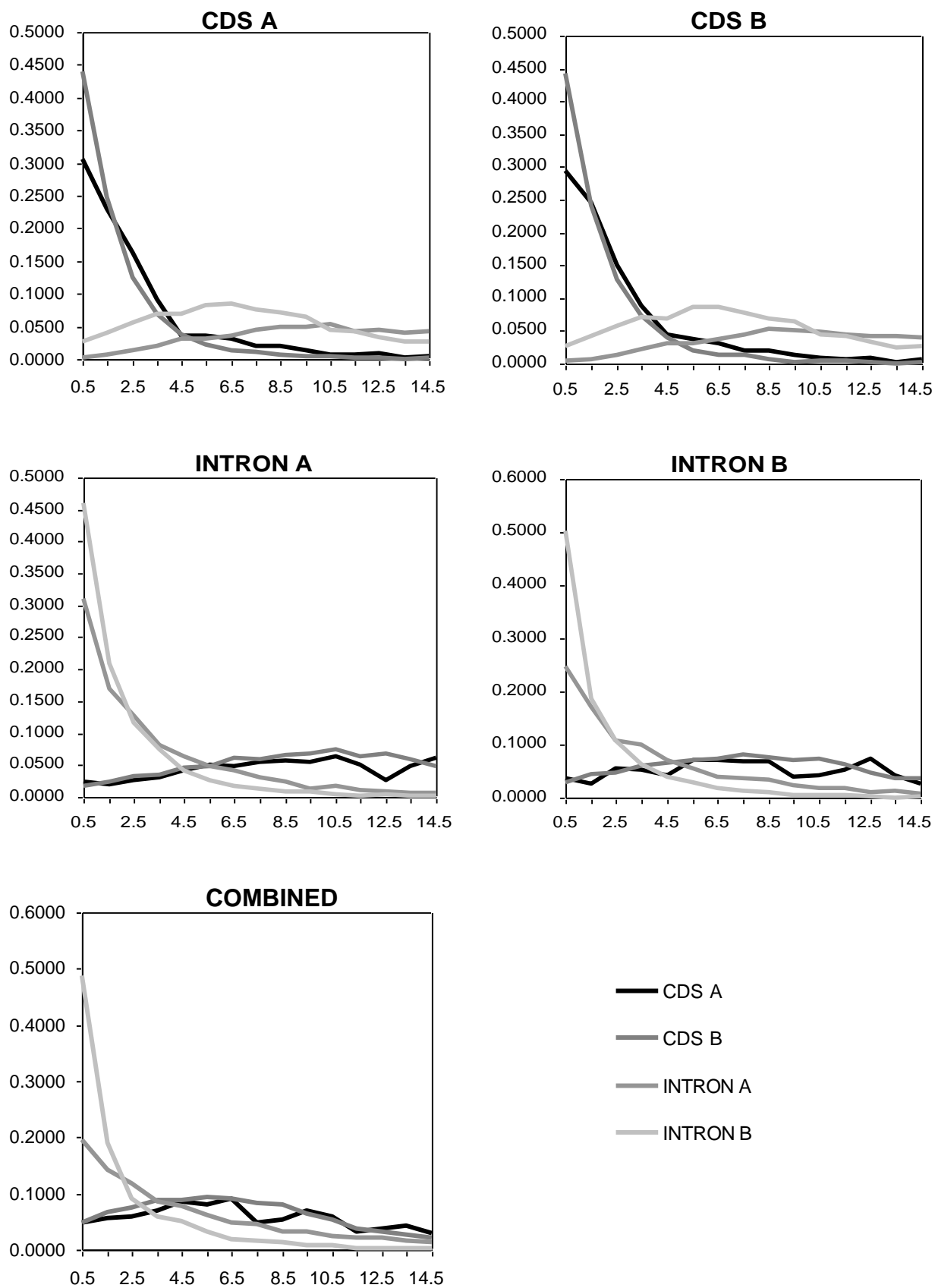


Fig 3: Chromosome 2R

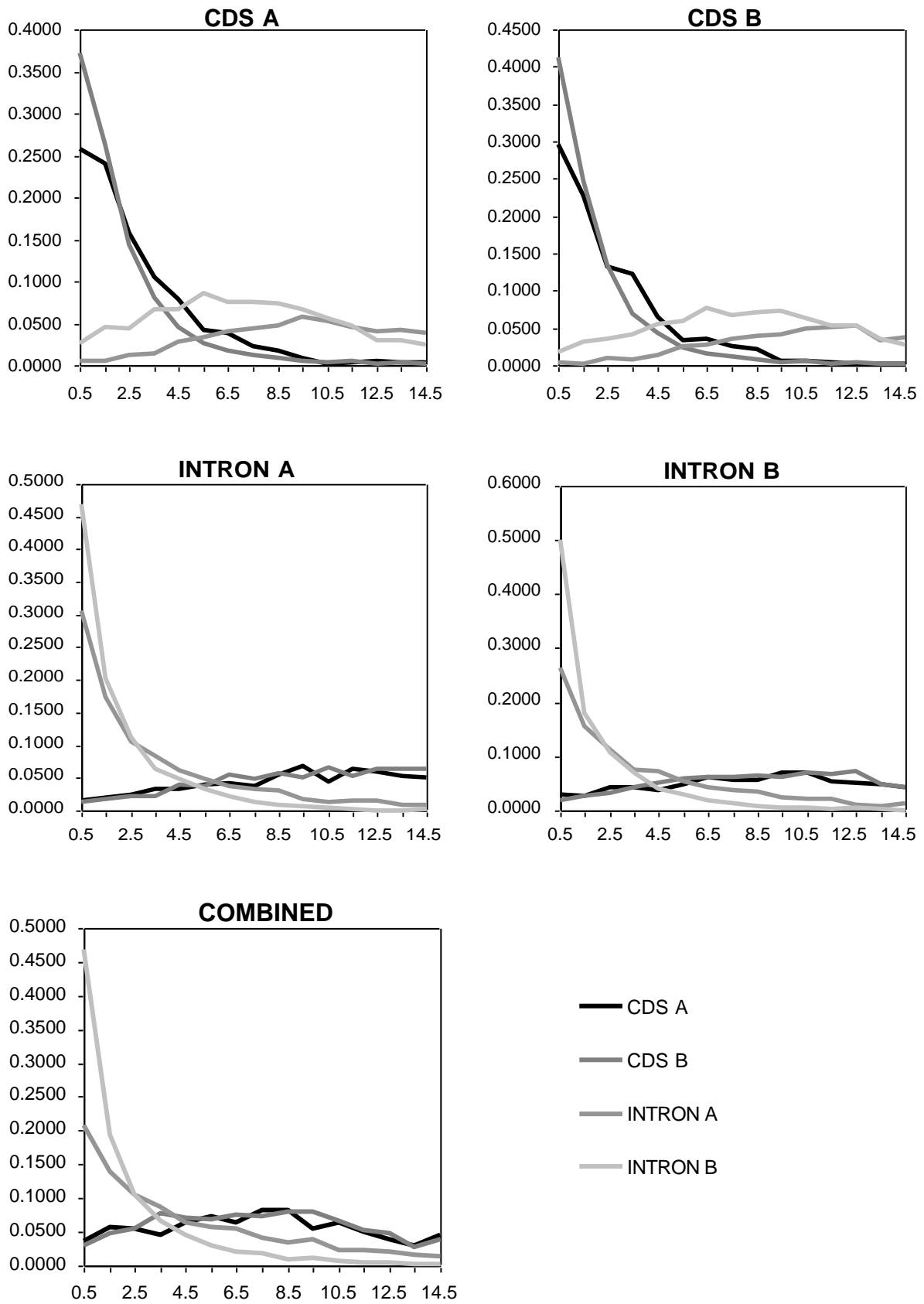


Fig 4: Chromosome 3L



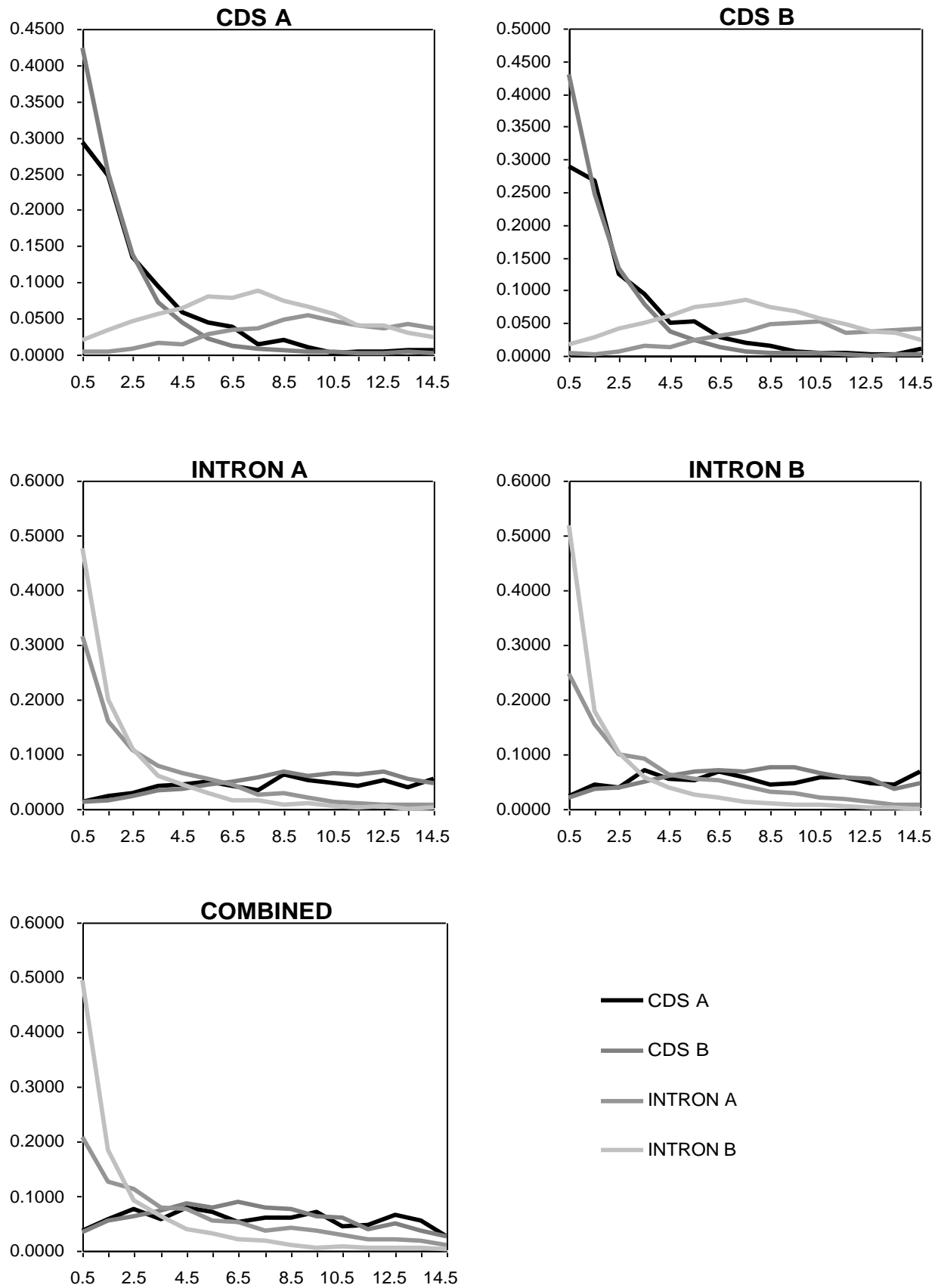


Fig5: Chromosome 3R

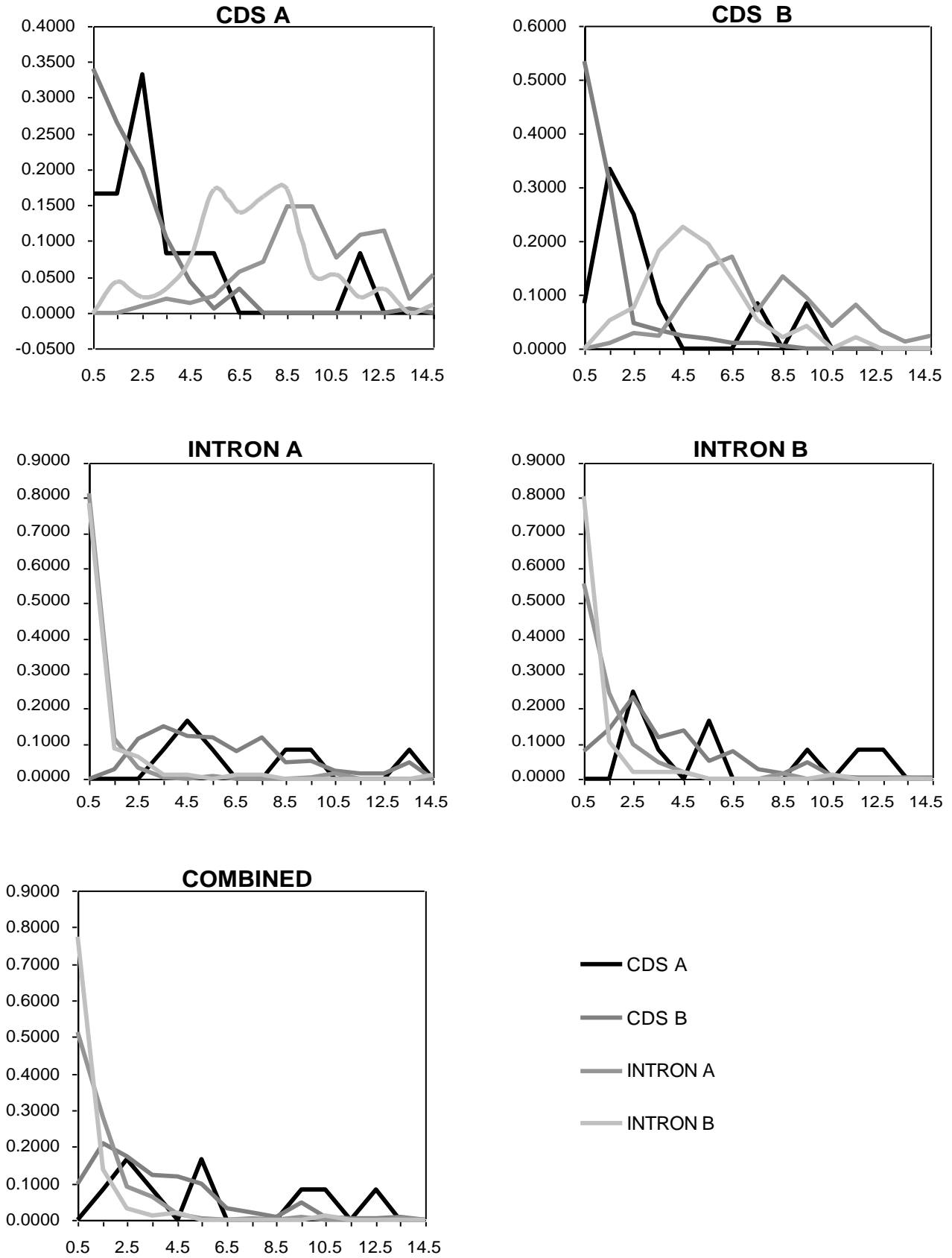


Fig 6: Chromosome 4

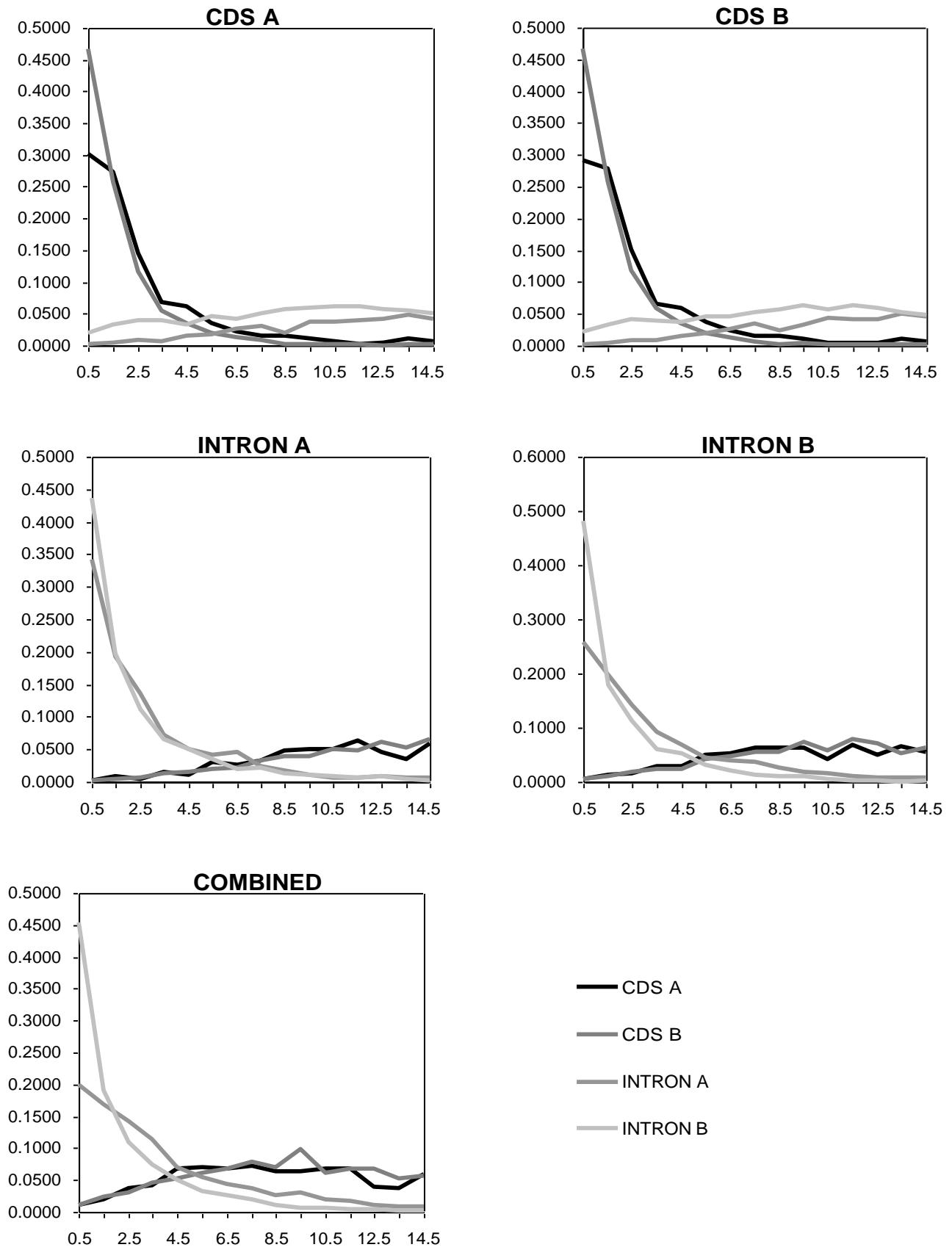


Fig 7: Chromosome X

## 7. OBSERVATIONS

We make the following observations from the Figures 2-7:

1. There is a difference between coding and non-coding sequences as expected.
2. An interesting point is that the two coding and the two non-coding types show difference among themselves.
3. The CDS A graph has the most fluctuation among all the graphs.
4. The behaviour is similar for all the chromosomes, except for Chromosome 4 which is slightly different from the others.
5. The best difference is expected by using  $C_5$  (combined), which is supposed to be a balance. However while using INTRON A as base (i.e., case  $C_3$ ) is the best for all chromosomes, though it is very similar and marginally better than  $C_5$ .
6. CDS A as base is also effective for differentiating coding and non-coding types, but matter is slightly complicated as it also creates a sizeable difference between INTRON A and INTRON B.

**N.B.** For each chromosome there are five charts. The heading of a chart relates to the case  $C_1$  to  $C_5$  respectively, while the four graphs relate to the four sequence types.

Final target is to determine a critical value  $M$  such that

- (a) Percentage of coding sequences with  $d^2 < M$  is high and Percentage of non -coding sequences with  $d^2 > M$  is low, or
- (b) Percentage of coding sequences with  $d^2 < M$  is low and Percentage of non -coding sequences with  $d^2 > M$  is high.

For this compare the cumulative relative frequencies for coding and non-coding and look for the maximum difference. This is done for each chromosome for each of the five cases  $C_1$  to  $C_5$ . The results are shown in Table5.

In the following table if overlook the chromosome 4 data then from the remaining it may easily decide on **4.5** as the critical value. Actually the critical value is  $d^2 = 4.5/500=0.009$ .Chromosome 4 gives always some abnormal result because of much fewer sequences than all the other chromosomes. If we consider the CDS A as the base then it may be concluded that almost 90% of coding sequences are below the critical point and also 90% of non-coding sequences are above that point..One more interesting result is that chromosome 2L and 3L gives almost same data leading to the same critical value.

Table 5: Critical value of  $500d^2$  for differentiating coding from non-coding sequences

C H R	CDS A (C <sub>1</sub> )		CDS B (C <sub>2</sub> )		INTRON A (C <sub>3</sub> )		INTRON B (C <sub>4</sub> )		COMBINED(C <sub>5</sub> )	
	CODING	NON-CODING	CODING	NON-CODING	CODING	NON-CODING	CODING	NON-CODING	CODING	NON-CODING
	CRITICAL VALUE		CRITICAL VALUE		CRITICAL VALUE		CRITICAL VALUE		CRITICAL VALUE	
2	0.8896	0.1307	0.8865	0.1211	0.1777	0.8636	0.2127	0.7891	0.2290	0.6822
	4.5		4.5		5.5		4.5		3.5	
2	0.8671	0.1098	0.8650	0.1143	0.1519	0.8216	0.1775	0.7309	0.1864	0.5983
	3.5		3.5		4.5		3.5		2.5	
3	0.8905	0.1510	0.8932	0.1068	0.1522	0.8495	0.1736	0.7814	0.2051	0.6770
	4.5		4.5		5.5		4.5		3.5	
3	0.9132	0.1245	0.9107	0.1140	0.1259	0.8050	0.1521	0.7151	0.2223	0.6672
	4.5		4.5		4.5		3.5		3.5	
4	0.9457	0.0826	0.8778	0.0663	0.0270	0.9116	0.2117	0.8360	0.2986	0.8329
	4.5		2.5		1.5		1.5		1.5	
X	0.9161	0.0980	0.9167	0.1007	0.0836	0.8952	0.0845	0.8158	0.1649	0.7806
	4.5		4.5		6.5		4.5		4.5	

## 8. CONCLUSION

The actual aim was to develop a method to classify a nucleotide sequence as coding or non-coding. The method developed can be summarized as follows:

*For the given sequence calculate  $d^2$  with CDS A as base. If  $d^2 < 0.009$  the sequence is classified as coding and if  $d^2 > 0.009$  it is classified as non-coding.*

Thus it is a powerful distinguishing criterion with 90% accuracy. Hope that the same type of criterion will apply to other species. It would be interesting to see if or how the critical value needs to be modified.

## References

- [1] Borodovsky, M., E. Koonin, and K. Rudd. (1994) New Genes in Old Sequences: A Strategy for Finding Genes in a Bacterial Genome. *Trends in Biochemical Sciences*,**19**:309-313.
- [2] Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. MolBiol*,**268**:78-94.
- [3] Crick F.H.C. and Watson J.D. (1954)The complementary structure of deoxyribonucleic acid. *Proc. Roy. Soc. (A)* **223**: 8096.
- [4] Fickett, J. W. (1982) Recognition of protein coding regions in DNA sequences, *Nucleic Acids Res*,**10**:5303-5318.
- [5] Mathe, C.,Sagot, MF, Schiex, T. , Rouze, P. (2002) *Current Methods of Gene Prediction, their Strengths and Weaknesses*, Nucleic Acids Research, **30**: 4103-4117.
- [6] Pachter, L. (1998) *Introduction to Computational Molecular Biology – Gene Recognition*: MIT Lecture notes.
- [7] Rao, C. R. (1952) *Advanced Statistical Methods in Biometric Research*. Wiley.
- [8] Pareek, C.S., Smoczynski, R. &Tretyn, A. Sequencing technologies and genome sequencing. *J Appl Genetics* **52**, 413–435 (2011). <https://doi.org/10.1007/s13353-011-0057-x>
- [9] Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research. *Nature* 422:835-847
- [10] Bansal V (2010) A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* 26:i318–i324



**Shashi Bajaj Mukherjee**, Assistant Professor Dr. B C Roy Engineering College, Durgapur. Ph.D from Jadavpur University. Working in the field of Bio-Statistics, Numerical Simulation, Soft Computing and Operations Research.



**Pradip Kumar Sen**, Professor Jadavpur University, M.Stat from ISI Kolkata, Ph.D from University of Chicago. Proficiency in German, French and Hindi language. Working in the field of Bio-Statistics, Statistics and Numerical Simulation.



## EFFECT OF ZYCOTHERM ON PROPERTIES OF POROUS FRICTION COURSE MIXES

Arijit Kumar Banerji<sup>1\*</sup>, Md. Hamjala Alam<sup>2</sup>, Antu Das<sup>3</sup>

<sup>1</sup>Department of Civil Engineering, Dr. B. C. Roy Engineering College, Durgapur, WestBengal, India. Email: [arijit.banerji@bcrec.ac.in](mailto:arijit.banerji@bcrec.ac.in),

<sup>2</sup>Department of Civil Engineering, Dr. B. C. Roy Engineering College, Durgapur, WestBengal, India. Email: [mdhamjala.alam@bcrec.ac.in](mailto:mdhamjala.alam@bcrec.ac.in)

<sup>3</sup>Quality Assurance and Control Engineer, B.S. GeoTech Pvt. Ltd., Konnagar, WestBengal, India., Email: [antudas94@gmail.com](mailto:antudas94@gmail.com)

**Abstract** — Warm mix asphalt (WMA) is a promising green technology with the potential to replace hot mix asphalt (HMA) in road construction. The purpose of this research is to assess the efficacy of Zycotherm WMA green technology in the construction of porous pavement. Porous pavement mixes comprised of open graded porous friction course materials with large voids reduce pressure on urban drainage systems and enhance natural water supply during rainstorms. Using WMA additive Zycotherm and VG 30 grade bitumen, this research focused on the properties of PFC grading bituminous mixes specified by the Indian Road Congress code IRC: 129-2019. Zycotherm is an eco-friendly nano-organic anti stripping additive that allows for lower mixing temperatures. Volumetric properties, water drain-down, permeability, and un-aged and aged abrasion resistance tests were performed on porous paving mixes with and without Zycotherm. This study examines PFC mixtures with bitumen contents of 5.0 to 6.0 percent. The experimental results were analysed in order to determine the most vital controlling parameters. According to the results, Zycotherm shows the potential to improve the performance properties of porous mixes.

**Keywords** — *warm mix asphalt, hot mix asphalt, Zycotherm, porous pavement.*

### 1. INTRODUCTION

Porous friction course mixes are a type of hot mix asphalt that is distinguished by a large percentage of inter-connected air voids, coarse granular grading, and suitable stone-on-stone contact. Bituminous PFC are also known as Open Graded Friction Courses (OGFC), Porous Asphalt (PA), Permeable Friction Courses (PFC), Open Graded Porous Asphalt (OGPA), Open Graded Asphalt Mixes (OGA) by various agencies around the world [1-3]. Between 19 and 50 mm thick, PFC layers have a minimum air void content of 18 % [4]. High air void content makes the PFC mixes permeable, which reduces the pressure on urban drainage systems and facilitates natural water supply in case of a rainstorm. Because of its porous structure, PFC can also be used as a storm water filter in addition to its other functions. PFC implements a uniform grading of aggregates in order to achieve the high proportion of air voids required. According to the research, PFCs provide a variety of benefits, including improved skid resistance, reduced pavement noise levels, and enhanced night time visibility during wet weather conditions, in addition to hydroplaning prevention [5-7]. The fact that many of these characteristics are in conflict with one another is obvious.

\*Corresponding author: Arijit Kumar Banerji, [arijit.banerji@brec.ac.in](mailto:arijit.banerji@brec.ac.in)

Unfortunately, the load bearing capacity of such PFC pavements is less than conventional pavements due to the absence of the fine particles [8]. It also weakens the adhesion bond between the binder and aggregate surface, resulting in an increased possibility of cracking, ravelling, and pothole formation. The moisture damage properties of PFC mixtures are crucial since they are expected to perform hydrological activities and are in continual exposure to water. In recent years, a significant number of porous asphalts have been developed with the goal of balancing these various requirements. In addition to providing a number of other advantages, the aggregate gradation of a bituminous PFC mixture has an impact on the ability of the mixture to withstand traffic loading. The effects of aggregate gradation on the properties of PFC were investigated for this study in order to evaluate volumetric and performance characteristics.

WMA is a recognised pavement construction technology that addresses environmental and economic concerns. In the road construction sector, WMA has the potential to substitute the practise of HMA [9]. The goal of WMA is to manufacture paving mixes at temperatures lower (20–40°C) than those used in hot mix. WMA is a technology that aims to minimise the amount of energy used and the amount of emissions produced during asphalt paving. The application of WMA technologies in asphalt paving can reduce energy usage by 40%, resulting in a reduction in greenhouse gas emissions [10]. Workers in the manufacturing plant and on the worksite are inhaling much fewer plumes of smoke. This also enables extended hauling distances, minimises the possibility of compaction difficulties, and necessitates less time for the compacted mixtures to cool before they are opened to traffic [11]. Currently, WMA technologies are classified into three broad categories: (i) foaming techniques, (ii) organic additive use, and (iii) chemical additive. In recent years, several articles on WMA technologies have discussed the usage of various additives. To test the resistance of WMA porous mix with Sasobit wax additive, Hamzah et al. used indirect tensile strength ratio test [12]. The results showed that dynamic water movement considerably reduced stripping resistance. Wurst and Putman compared Evotherm and foamed WMA with HMA in a warm-mix environment [13]. They discovered that by employing the WMA technologies, fibres from OGFC mixes can be removed when comparing three major criteria: drain down, permeability, and abrasion resistance. According to Bairgi et al., compaction and long-term ageing both contribute to the promotion of rutting and stripping parameters [14]. WMA and OGFC mixtures comprising RAP and PMB was analyzed by Ingrassia et al. [15] using an in-situ falling weight deflectometer. They found that the OGFC and WMA combinations outperformed the HMA mixture in terms of reliability and durability. However, the differences in aggregate gradations, bitumen sources, mix designs, and testing conditions in India, in comparison to other countries, require specific studies to gain an understanding of the behaviour of WMA mixes. Warm mix PFC has been evaluated in limited performance trials with WMA technologies in recent years. Furthermore, because of the above WMA properties, ravelling and drain down could be reduced by using warm mix PFC, especially when anti-stripping WMA additives are utilized.

## 2. OBJECTIVES AND SCOPE

The purpose of this study is to assess the practicality of the application of the ZycoTherm warm mix additive to PFC mixes. PFCs have not yet been tried by Indian authorities in national highways or expressways, but recently 2019 specifications have been recommended.



PFC mixes with gradations suggested by the Indian Roads Congress (IRC: 129-2019) [16] specifications for open-graded friction courses were explored in this study. The performance and reliability characteristics of PFC combinations were studied by taking into account the following mix design qualities: volumetric properties, porosity, drain down, un-aged and aged abrasion loss. The Marshall approach is used to compact the mix, which will consist of 25 blows on each side of the mixture. The experimental data was analysed to determine the most significant influencing parameters and their relative importance.

### 3. MATERIAL DESCRIPTION AND MIX PREPARATION

#### A. Materials Used

The investigation focused on the PFC mixes that corresponded to the gradation specified by the IRC: 129-2019 requirements, as shown in Table 1. A grading specification that was closer to the middle of the range was chosen [17]. Portland Pozzolana cement was employed as the filler fines. In accordance with Indian Standards and ASTM test procedures, the physical properties of aggregate and bitumen were determined. Detailed results of the tests are reported in Table 2. Additionally, each mixture includes WMA ZycoTherm at 0.1% by weight of bitumen. It is an innovative, high performance anti stripping additive for asphalt mixes that, in comparison with conventional anti stripping additives, chemically bonds with the aggregates to generate a flawless barrier that lasts for an extended period of time. It enables 100% coating and wetting of the pavement, as well as improved anti-stripping capabilities, resulting in increased moisture resistance of the pavement [18,19]. When compared to conventional additives, ZycoTherm modified bitumen coats aggregate surfaces substantially faster and with less effort than other types of bitumen. Its addition has the effect of lowering the temperature at which asphalt mixtures are produced and compacted. ZycoTherm is the chemical WMA additive with the following properties given in the Table 3 [18]. The bitumen was first heated to 140°C for mixing.

Table 1 - IRC: 129-2019 Aggregate Grading

IS Sieve Size (mm)	13.2	9.5	4.75	2.36	0.075
% Passing	100	92.5	30	7.5	3

Table 2 – Physical Properties of Coarse Aggregate and Bitumen

Materials	Properties	Results	Standards
Aggregates	Aggregate Impact Value, %	12.28	IS 2386 (P4)
	Combined Flakiness and Elongation Index, %	24.67	IS 2386 (P1)
	Fractured Particles, %	97	ASTM D8521
	Water absorption, %	0.55	ASTM C127
	Specific gravity (bulk)	2.82	IS 2386 (P3)
	Specific gravity (apparent)	2.85	
Bitumen	Los Angles Abrasion, %	19.34	ASTM C131
	Stripping Test	100	
	Specific gravity at 27°C	1.02	IS 1202
	Penetration at 25°C, 100 g, 5 s	57	IS 1203

Softening point, (R&B), °C	48.5	IS 1205
Ductility at 27°C, cm	>100	IS 1208
Flash Point, °C	310	IS 1209

Table 3 - Physical Properties of WMA Additive ZycoTherm [18]

Property	Descriptions
Specific gravity	0.97
Viscosity	100-500 cps
Flash point	>80°C
Colour	Pale yellow
pH value	10% soluble in neutral water
Physical state	Liquid

### B. Bituminous PFC Mix Preparations

Asphalt mixes are made by first preheating aggregates at mixing temperature in oven before mixing them together. The compaction temperature was maintained in the oven for one hour after mixing. For HMA mixes, 165°C mixing temperatures was used, and for WMA mixes, 145°C mixing temperatures was adopted. Compaction was performed at a temperature that was 10 degrees Celsius lower than mixing. All of the PFC mixtures were made into standard specimens of 100 mm diameter by 63.5 mm height. 5.0 percent, 5.50 percent, and 6.0 percent bitumen were used in the design trials. 25 blows of the Marshall Hammer was applied to each side, and the samples were let cool for 24 hours at room temperature before being investigated in accordance with IRC 129: 2019 [16].

## 4. PFC PERFORMANCE TESTS

### A. Volumetric Properties of Compacted Specimens

The service life of PFC pavements varies depending on the environment, vehicular traffic and loads, as well as the construction and design standards used to construct them. The minimal bitumen content at which the PFC mixes are capable of meeting ASTM D7064 [20] design standards was chosen as the optimum bitumen content (OBC). The volumetric properties were calculated using the bulk specific gravity and maximum specific gravity results. The theoretical maximum density of the loose mix was determined using effective specific gravity as per standards of ASTM D 2041 [21]. When the maximum specific gravity and bulk specific gravity were obtained, the air void was calculated by multiplying the two quantities. Percentage of voids in compacted mix ( $VCA_{mix}$ ) and percentage of voids in dry-rodded state ( $VCA_{drc}$ ) were used to detect whether there was stone-on-stone (SoS) contact in the PFC mix gradations. The density of a specific coarse aggregate sample is determined by conducting a dry rodded test. Using a constant volume cylinder filled with three layers and 25 tamps per layer, this test provides an understanding of the internal arrangement of aggregates in a dry state. PFC mix SoS contact condition was approved by examining the  $VCA_{mix}$  to  $VCA_{drc}$

ratio. All of the mixtures that corresponded to the gradation were observed to have mean void percentage values greater than 18 percent. Table 4 shows the results of these studies for maximum specific gravity, bulk specific gravity of PFC mix,  $VCA_{mix}/VCA_{drc}$ , and void percentage for the purposes of comparison.

Table 4 - Volumetric Properties of PFC mix

Property	Results		
BC (%)	5.0	5.5	6.0
Air Void, %	21.79	21.44	19.58
$VCA_{mix}$	37.4	37.6	36.5
$VCA_{drc}$	44.1		
$VCA_{mix}/VCA_{drc}$	0.84	0.85	0.82

## B. Permeability

Permeability or hydraulic conductivity, refers to the rate at which water passes through a material. Drainage properties are critical to the performance of a PFC mix. Pore size measurements help determine the drainage characteristics of these mixes. The contact surface of the Marshall sample with mould was double coated to avoid water leakage. The time it took for the water to descend from initial head ( $h_1$ ) to final head ( $h_2$ ) was recorded in seconds. Equation 1 was used to compute the permeability coefficient ( $k$ ) and the results ranged from 92 to 140 m/day. A permeability rate of 100 m/day was advised by Mallick and Chen to provide adequate performance [22,23]. Because permeability coefficient value should be greater than 100 metres per day for proper drainage, this mix's bitumen concentrations of 5.0 and 5.5 percent met this requirement. Permeability was shown to be closely linked to air-void trends, which is generally believed to be true.

$$(1) \quad K(m/day) = \frac{(3600 \times 24)}{1000} \frac{al}{At} \ln \frac{h_1}{h_2}$$

where;  $l$  is sample thickness in mm,  $a$  is standpipe cross-sectional area in  $mm^2$ ,  $A$  is sample cross-sectional area in  $mm^2$  and  $K$  is permeability constant. Table 5 summarises the permeability data for each of the PFC combinations.

Table 5 - Permeability and Air Void results

Property	Results		
BC (%)	5.0	5.5	6.0
K(m/day)	140	133	92
Air Void (%)	21.79	21.44	19.58
K(m/day)	Adding 0.1% ZycoTherm@5.5%: 178		
Air Void (%)	21.82		

### C. Cantabro Abrasion Loss

The compacted specimens were subjected to the Cantabro abrasion test in terms of determining their durability. In the un-aged state, the abrasion loss should not exceed 20%, and in the aged state, it should not exceed 30%. For the mix design of PFCs, most agencies now suggest this test as either a mandatory or optional test. The abrasion loss on un-aged and aged samples are denoted as UAL and AAL, respectively.



Figure 1 - (a) Specimen in abrasion cylinder (b) Fully compacted specimen and after abrasion.

The un-aged abrasion loss test was performed on a Los Angeles abrasion drum (Figure 1) at 30 to 35 revolutions per minute for 300 revolutions. Samples of asphalt mixtures were tested by determining how much weight were reduced when compared with their initial weight. As a result, the bitumen in PFC hardens faster resulting in a loss of cohesive and adhesive strength, which can lead to ravelling. Consequently, an accelerated ageing test was carried out on the mix design to see how it performs at 85°C for 120 hours. The test is performed according ASTM D7064 [20] in both aged and un-aged conditions. Tests will be performed on the specimens after they have been cooled to 25°C and stored for four hours. Individual sample abrasion loss should not above 50 percent, whereas UAL and AAL sample abrasion loss should not exceed 20 percent as well as 30 percent, respectively. The oxidation of bitumen makes it more stiff, which makes the bond between aggregate particles and the bitumen more brittle, which makes abrasion loss more likely in aged samples. The variation of UAL and AAL of HMA and WMA mix are shown in Table 6.

Table 6 - Unaged Abrasion Loss and Aged abrasion Loss results

Property	Results			
BC (%)	5.0	5.5	WMA	6.0
Unaged Abrasion Loss (%)	16.81	13.78	22.51	13.45
Aged Abrasion Loss (%)	10.85	10.70	14.55	8.92

### D. Drain Down Test

Draindown testing was performed on un-compacted specimens in accordance with AASHTO T305 [24] and ASTM D6390 [25] standards. The test allows for the evaluation of an asphalt mixture's drain down potential at the stages of mixture design and field manufacturing, respectively. A thin bitumen coating is left behind after draindown, which is insufficient to prevent particles from being dislodged by traffic. It is also possible for the thin bitumen film to age more quickly and become brittle, thus worsening the ravelling problem. Increased bitumen content in PFC mixtures will have a greater chance of causing bitumen to drain away during transportation. This will result in dispersion of bitumen and improper bituminous mix, which will cause ravelling to occur.



Figure 2 - (a) Specimen conditioning in oven (b) Flask with drained material.

The Schellenberg Binder Drainage Test was used to measure the amount of bitumen drained in this investigation. This experiment used glass jars with a diameter of 98mm and a height of 136mm (Figure 2). The sample is kept in glass jars for an hour at a temperature 15°C above mixing and compaction. After a 60 minutes, the jar is tipped to measure the remaining material. The Drain Down value should be within 0.3 percent of each other when calculated using Equation 2. Table 7 shows the drain down values for different bitumen contents.

$$\text{Drain down (\%)} = \frac{(\text{Final Weight of flask with Drain sample} - \text{Initial Weight of flask})}{\text{Initial Weight of flask}} \times 100 \quad (2)$$

Table 7 - Drain down Test results

Property	Results			
BC (%)	5.0	5.5	WMA	6.0
Drain down (%)	0.20	0.21	0.19	0.24

## 4. CONCLUSION

The key objective of this study was to assess the practicability of using WMA technologies to produce acceptable PFC mixtures. This study examined ZycoTherm WMA mixes with HMA PFC based on draindown, permeability, and abrasion resistance as the key criteria. There were two asphalt mixtures examined and tested: one with 0.1 percent ZycoTherm, and the other with VG bitumen (HMA mix). For this experiment, the mixing and compacting temperatures were varied so that the performance of the mixes could be evaluated. First, volume and physical measurements were taken, followed by Cantabro, drain down, and permeability tests.

The stone-on-stone contact criterion was met by this gradation. They have good volumetric and permeability qualities for PFC mixtures compacted with 25 blows. Compared to PFC samples, WMA samples have higher water permeability ratings because they contain more air. Mixes with a bitumen concentration of 5.0 and 5.5 percent had air voids content of approximately 20 percent. Both the Unaged Abrasion Loss and Aged abrasion Loss values lies inside the 20% and 30% thresholds, respectively. Traditional PFC specimens are slightly more susceptible to compaction than ZycoTherm specimens, perhaps due to the lower manufacturing temperature. Mixes treated with WMA ZycoTherm reduced drain down losses to less than 0.20%. According to these data, WMA technology may be used to lay porous asphalt at a lower temperature without a substantial deterioration. However, further research is required to better understand these mixtures.

**Acknowledgement:** The research is based upon the work supported by Dr. B. C. Roy Engineering College, Durgapur.

## References

- [1] Shankar, A. U., Suresha, S. N., & Saikumar, G. M. V. S., "Properties of porous friction course mixes for flexible pavements", *Indian Highways*, 42(3), 2014.
- [2] Suresha, S.N, Varghese George and Ravi Shankar, A.U., "Investigation of Porous Friction Courses (PFC) and Mixes: a Brief Overview", *Indian Highways*, Vol. 35, No.7, p. 21-43, 2007.
- [3] Banerji, A. K., Das, A., Mondal, A., Obaidullah, M., Biswas, R., & Saha, U., "Performance Assessment of Porous Friction Course (PFC) Mix modified with Cement as Filler Material", *American International Journal of Research in Science, Technology, Engineering & Mathematics*, 15-532, p. 83-88, 2015.
- [4] Mansour, T. N., & Putman, B. J., "Influence of aggregate gradation on the performance properties of porous asphalt mixtures", *Journal of Materials in Civil Engineering*, 25(2), 281-288, 2013.
- [5] Huber, G., "Performance Survey on Open-Graded Friction Course Mixes." *Synthesis of Highway Practice 284*, National Cooperative Highway Research Program, Transportation Research Board, 2000.
- [6] Kandhal, P. S. "Design, construction and maintenance of open-graded asphalt friction courses", *National asphalt pavement association information series 115*. America.
- [7] Alvarez, A. E., Martin, A. E., & Estakhri, C. "A review of mix design and evaluation research for permeable friction course mixtures", *Construction and Building Materials*, 25(3), 1159-1166, 2011.
- [8] Kandhal, P. S., & Mallick, R. B. "Open-Graded Asphalt Friction Course: State of the Practice. National Center for Asphalt Technology Report", (98-7), 1998.
- [9] Prowell, B. D., Hurley, G. C., & Crews, E., "Field performance of warm-mix asphalt at the NCAT test track", 86th Annual Meeting of the Transportation Research Board, Transportation Research Board, 2007.
- [10] Vaitkus, A., Čygas, D., Laurinavičius, A., & Perveneckas, Z., "Analysis and evaluation of possibilities for the use of warm mix asphalt in Lithuania", *Baltic J. Road Bridge Eng.*, 4(2), 80-86, 2009.
- [11] M. Zaumanis, *Warm Mix Asphalt Investigation Master of Science Thesis*, Technical University of Denmark, 2010.

- [12] M.O. Hamzah, M.Y. Aman, Z. Shahadan, M.H.M. Rosli, "Laboratory Evaluation of the Dynamic Stripping Test On Porous Asphalt Incorporating Sasobit", 7th International Conference on Road and Airfield Pavement Technology, Bangkok, 2011, p. 130-141.
- [13] Wurst III, J. E., & Putman, B. J., "Laboratory evaluation of warm-mix open graded friction course mixtures", Journal of Materials in Civil Engineering, 25(3), 403-410, 2013.
- [14] Bairgi, B. K., Tarefder, R. A., & Ahmed, M. U., "Long-term rutting and stripping characteristics of foamed warm-mix asphalt (WMA) through laboratory and field investigation", Construction and Building Materials, 170, 790-800, 2018.
- [15] Ingrassia, L. P., Cardone, F., Ferrotti, G., & Canestrari, F., "Monitoring the evolution of the structural properties of warm recycled pavements with Falling Weight Deflectometer and laboratory tests", Road Materials and Pavement Design, S69-S82, 2021.
- [16] IRC: 129 Specifications for Open-Graded Friction Course. IRC, New Delhi, India, 2019.
- [17] Banerji, A. K., Das, A., Mondal, A., Biswas, R., & Obaidullah, M., "A Comparative evaluation on the properties of HMA with variations in aggregate gradation of laboratory and field produced mixes. Int. J. Eng. Res. Technology., p. 441-448, 2014.
- [18] Zydex, ZYCOTHERM Material Safety Data Sheet. Gotri, Vadodara:Zydex, Editor, 2014.
- [19] Tripathi, P., Ray, D. S., "Optimization of Different Anti Stripping Agents in Construction of Flexible Pavement:' International Journal of Engineering Research & Technology, 9(7), p. 670-681, 2020.
- [20] ASTM D 7064, Standard practice for open graded friction course (OGFC) mix design, ASTM International, ASTM International, West Conshohocken, PA, 2013.
- [21] ASTM D 2041, Standard test method for theoretical maximum specific gravity and density of bituminous paving mixtures, ASTM International, West Conshohocken, PA, 2011.
- [22] Mallick, R.B., Kandhal, P.S., Cooley, L.A. Jr., and Watson, D.E., "Design, Construction, and Performance of New-Generation Open-Graded Friction Courses", J. Association of Asphalt Paving Technologists, 69, 391-423, 2000.
- [23] Chen, J.S., Lin, K.Y., and Young, S.Y., "Effects of Crack Width and Permeability on Moisture-Induced Damage of Pavements", J. Materials in Civil Engineering, 16(3), 276-282, 2004.
- [24] AASHTO T305. Standard Method of Test for Determination of Draindown Characteristics in Uncompacted Asphalt Mixtures, American Association of State Highway and Transportation Officials, Washington, D.C., 2001.
- [25] ASTM D 6390. Standard Test Method for Determination of Draindown Characteristics in Uncompacted Asphalt Mixtures, West Conshohocken, P.A., 2005.



**Arijit Kumar Banerji** is a B.Tech in Civil Engineering from BIET Suri and an M.Tech in Transportation Systems Engineering from IIT Guwahati. Since July, 2014 he is working as Assistant Professor at Dr. B.C Roy Engineering College, Durgapur, West Bengal. Presently he is also pursuing his Phd at NIT, Durgapur as part-time research scholar. He is having more than seven years of teaching experience. His areas of interests are Pavement material characterization, Pavement Design and Finite Element Modelling of pavement.



**Md. Hamjala Alam** is a B.Tech in Civil Engineering from Narula Institute of Technology, Kolkata and an M.Tech in Environmental Science & Engineering from Indian School of Mines Dhanbad. Since August, 2014 he is working as Assistant Professor in Civil Engineering at Dr. B.C Roy Engineering College, Durgapur, West Bengal. His areas of interests are Water Quality, Air and Noise Pollution and Control, Soil Stabilization, Soil Strength enhancement, Concrete properties and Pavement Design.



**Antu Das** is a B.Tech in Civil Engineering from Dr. B. C. Roy Engineering College, Durgapur and currently working as Quality Assurance and Control Engineer in B.S. GeoTech Pvt. Ltd., Konnagar, WestBengal. He has published two international journal paper and his areas of interests are highway design, soil investigation and behaviour of structures.



# Electrical Conduction Phenomenain Some Promising Lithium Ion Doped Glassy Ceramics

Aditi Sengupta<sup>1</sup>, Dr. Sanjib Bhattacharya<sup>2</sup>, Dr. Chandan Kr. Ghosh<sup>3</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Siliguri Institute of Technology, Darjeeling-734009, West Bengal, India  
email: [aditi\\_sengupta1987@yahoo.com](mailto:aditi_sengupta1987@yahoo.com)

<sup>2</sup>Composite Materials Research Laboratory,  
Department of Physics, University of North Bengal,  
UGC-HRDC, University of North Bengal, District: Darjeeling, Pin-734013, West Bengal, India  
email: [sanjib\\_ssp@yahoo.co.in](mailto:sanjib_ssp@yahoo.co.in); [ddirhrdc@nbu.ac.in](mailto:ddirhrdc@nbu.ac.in)

<sup>3</sup>Department of Electronics and Communication Engineering, Dr. B.C. Roy Engineering College, Durgapur-713026, West Bengal, India  
email: [chandan.ghosh@bcrec.ac.in](mailto:chandan.ghosh@bcrec.ac.in)

**Abstract**—A glass ceramic sample,  $0.1\text{Li}_2\text{O}-(0.9)(0.5\text{MoO}_3-0.4\text{V}_2\text{O}_5-0.1\text{ZnO})$  has been prepared using a melt quenching route. The ionic conductivity, temperature and frequency dependence has been established following Almond-West formalism and Jonscher's universal power-law respectively. The Study of its dc conductivity ( $\sigma_{dc}$ ) shows the thermally activated nature of the sample. The variation of activation energy and hopping frequency values with composition corresponding to AC conductivity (EAC) has been established. This is the cause behind The frequency independent conductivity (plateau region) in the low-frequency zone formed due to the sub diffusive motion of  $\text{Li}^+$  ions. Whereas, the correlated and pseudo-three-dimensional motion of  $\text{Li}^+$  ions in percolating networks causes conductivity in high-frequency dispersive regions.

**Keywords** —  $\text{Li}^+$  ion migration; Glass-ceramics;; Ionic conductivity; Hopping Frequency.

## I. INTRODUCTION:

Structural and electrical properties of nanophase materials are of great scientific interest in recent times <sup>[1]</sup>. The main reason behind this interest is the improved electrical properties of these nano phased materials, over its bulk counterparts <sup>[1]</sup>. Scientific efforts are being made towards the evolution of better materials suitable for engineering practice, after comparing the electrical and mechanical properties of so far available conventional materials <sup>[1]</sup>. The present work conveys the development of a lithium based glassy system using conventional melt quenching route and interpretation of its electrical transport properties such as activation energy, ionic conduction, hopping frequency etc. Lithium is one of the key components of rechargeable batteries and supercapacitors in recent era with the promise of renewable energy storage <sup>[2-4]</sup>. Rechargeable batteries are the most useful part of most of today's human needs such as electric vehicles,



smart phones, computers etc [2-4]. The responsibility for the product's zero expansion makes lithium indispensable to every glass-ceramic, which further ensures its use in high temperature ranges without voltage breakage [5]. Incorporation of suitable metal ions such as molybdenum and vanadium, improves the electrical properties, chemical inertness and thermal stability of the glassy system [6].

To shed some light on conductivity spectra and electrical relaxation mechanisms in lithium-ion conductors, Jonscher's power-law model [7, 8] can be employed. In this law [7, 8], total conductivity of lithium-ion conductor can be described as:

$$\sigma(\omega) = \sigma_0 + A\omega^S \quad (1)$$

where, A is the pre factor, S is the frequency exponent and  $\sigma_0$  is the low frequency or DC conductivity. The conductivity spectra of them can be analyzed at various temperatures using Almond-West formalism (power-law model) (9),

$$\sigma(\omega) = \sigma_{dc} [1 + (\omega / \omega_H)^n] \quad (2)$$

, which is the combination of the dc conductivity ( $\sigma_{dc}$ ), hopping frequency ( $\omega_H$ ) and a fractional power law exponent (n). To get sufficient information on the conducting system [7], this model has been used to fit the experimental conductivity data with respect to frequency.

Investigation of electrical transport properties [6] of Li<sub>2</sub>O doped glass ceramic is expected to reveal their structural information, which is essential for the knowledge of conduction mechanisms. In the present work, electrical conduction behaviour of Li<sub>2</sub>O doped new glass ceramic has been studied to shed some light on new features of lithium-ion conductors. The effect of Li<sub>2</sub>O doping in the glassy matrix is expected to be interesting not only for technological applications but also for academic interest.

## II. EXPERIMENTAL PROCEDURE:

A glass ceramic, xLi<sub>2</sub>O-(1-x) (0.5MoO<sub>3</sub>-0.4V<sub>2</sub>O<sub>5</sub>-0.1ZnO) for x = 0.1 has been prepared using melt quenching route from the precursor powders Li<sub>2</sub>O, MoO<sub>3</sub>, V<sub>2</sub>O<sub>5</sub> and ZnO. The mixtures are heated in an alumina crucible in an electric furnace in the temperature range 750 °C for 30 minutes and quenched between two aluminium plates at room temperature. Partially transparent glass ceramic of thickness ~ 1 mm has been formed. Schematic diagram of experimental setup for present work is presented in Fig. 1. The conducting silver paste has been painted on both sides of the glass ceramic flakes, which acts as electrodes and mounted into the two-probe system and placed in a furnace (Fig. 1). The temperature variation has been achieved using a dedicated temperature controller. The capacitance (C), conductance (G) and dielectric loss tangent (tanδ) of an as-prepared sample have been measured using a programmable high precision LCR meter (HIOKI, model no. 3532-50) at various temperatures in the frequency range 42 Hz - 5MHz. The microstructures of the as-prepared samples are explored by scanning electron microscopic (JEOL made) studies.

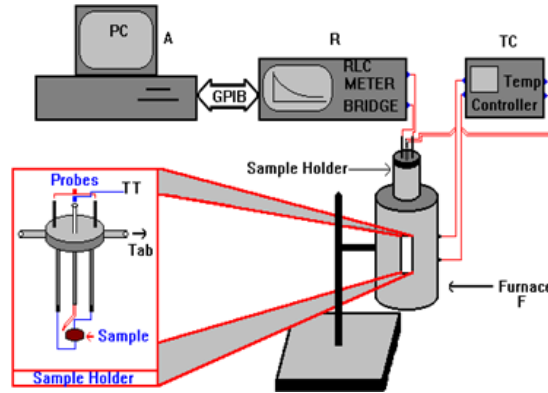


Fig 1: Experimental setup for electrical measurement

### III. RESULTS AND DISCUSSION:

DC electrical conductivity can be estimated from the complex impedance plots [8, 9]. Fig. 2 represents DC electrical conductivity with reciprocal temperature, which shows thermally activated nature.

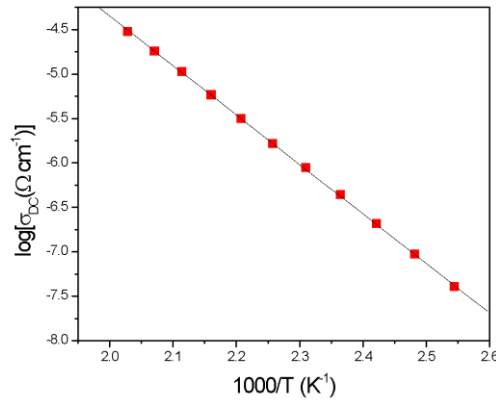


Fig 2: Temperature dependency of DC conductivity

DC conductivity may also be introduced as a function of temperature which follows Arrhenius relation:

$$\sigma_{DC} = \sigma_0 e^{-\frac{E_\sigma}{KT}} \quad \sigma_{DC} = \sigma_0 e^{-\frac{E_\sigma}{KT}} \quad (3)$$

where,  $E_\sigma$  is the activation energy,  $T$  is the absolute temperature (moderate) and  $K$  is the Boltzman constant. DC conductivity, estimated from its bulk resistance, has been analyzed using Eq. (3), which shows thermally activated nature. Activation energy ( $E_\sigma$ ) corresponding to DC conductivity has been estimated from the best fitted straight line fit and presented in Table-1.

TABLE-1: DIMENSIONALITY,  $E_{\sigma}$  AND  $E_H$  AT VARIOUS TEMPERATURES. (ESTIMATED ERRORS OF MEASUREMENT ARE ALSO INCLUDED)

Temperature (K)	Dimensionality (n) ( $\pm 0.05$ )	$E_{\sigma}$ (eV) ( $\pm 0.05$ )	$E_H$ (eV) ( $\pm 0.05$ )
403	1.02436	-1.10201	-1.07244
413	0.8301		
433	0.75831		
443	0.73485		
463	0.83682		
483	0.86349		
493	0.80629		

AC conductivity spectra <sup>[1, 7-9]</sup> have been analyzed at various temperatures using Eq.(2). The AC conductivity spectra at various temperatures have been presented in Fig. 3. It is evident from Fig. 2 that at low frequency, the conductivity becomes flat. This frequency independent conductivity corresponds to the DC conductivity. At higher frequencies, the AC conductivity shows dispersion and follows a power law nature as presented in Eq. (1). It is obvious from Fig. 3 that AC conductivity of present glass ceramic shows thermally activated nature and conductivity isotherm increases with temperature nature. Flat-conductivity at low frequency regime may be due to the diffusion motion of Li<sup>+</sup> ions <sup>[1]</sup>. The higher- frequency- dispersion may correspond to a nonrandom, correlated and sub-diffusive motion <sup>[1]</sup> of Li<sup>+</sup> ions, which may be anticipated for interionic interaction. It may result in uniform change in power law exponent.

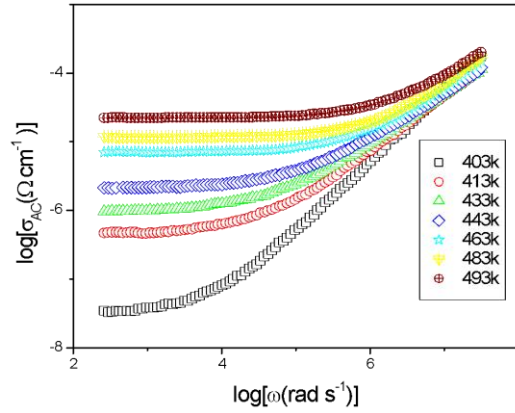


Fig 3: Temperature and frequency dependent AC conductivity spectra

Hopping frequency ( $\omega_H$ ) and a fractional power law exponent ( $n$ ) have been estimated from the fitting of experimental conductivity spectra using Eq. (2). It is found from Table-1 that frequency exponent  $n$  is less than unity at higher temperature and depends significantly on composition due to some differences in structure. But at lower temperature  $n$  is found to be greater than unity, which shows percolation type motion. As the temperature increases,  $\text{Li}^+$  ion motion follows three dimensional conduction pathways [10, 11]. Scientific works [10, 11] show that the power-law exponent values are the indicators of the dimensionality of conduction pathways. Power of frequency  $\sim 2/3$  [10, 11] may anticipate composition as well as temperature-independence nature of three dimensional motion of  $\text{Li}^+$  ions in the present system. Two-dimensional ion motion in  $\text{Na } \beta\text{-alumina}$  and one-dimensional ion motion in hollandite [11] indicate smaller exponents  $n=0.58\pm 0.05$  and  $n=0.3\pm 0.1$  respectively. So, AC conductivity is expected to be influenced by the dimensionality of the ion-conduction space, which is manifested by the frequency-exponent.

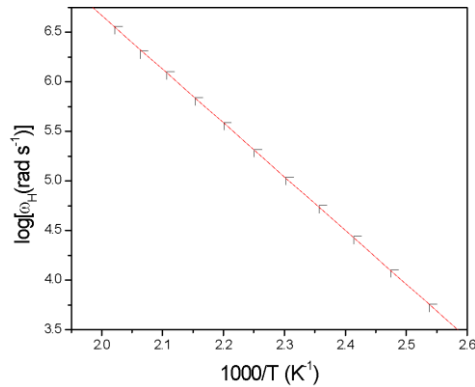


Fig 4: Temperature dependency of Hopping frequency

Fig. 4 represents temperature dependence of  $\omega_H$ , which exhibits thermally activated nature [8]. Activation energy corresponding to hopping frequency ( $E_H$ ) is estimated from slopes of the curve as presented in Fig. 4

and presented in Table-1. This result indicates that electrical conductivity is mainly controlled by hopping frequency in the present glass ceramics.

The AC conductivity at five frequencies with reciprocal temperature is presented in Fig. 5. Here, DC conductivity is included for comparison. It is observed that at lower temperatures the AC conductivity is spreading out due to the dominant nature of AC conductivity over the DC conductivity. As the temperature is increased, the AC conductivity approaches the DC conductivity. It is also noteworthy that the AC conductivity increases with the increase in frequency.

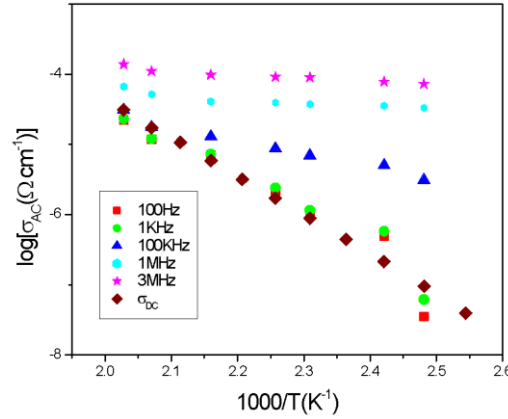


Fig 5: AC conductivity at five frequencies with reciprocal temperature

The conduction phenomenon can also be observed by the complex impedance plot shown in figure 6. In this plot both x and y axis represent complex impedances  $Z'$  and  $Z''$  respectively. The radius of curvature in different temperatures replicates the resistivity variation with temperature. The higher spread of semi-circular plots indicates lower conductivity (high resistance) which ensures the decrease of conductivity with increment of radius of curvature.

In the equivalent circuit (parallel RC), the overall AC impedance of the present circuit can be represented as<sup>[12]</sup>:

$$\begin{aligned} \frac{1}{z} &= \frac{1}{Z_R} + \frac{1}{Z_C} \\ &= \left( \frac{1}{R} + j\omega C \right)^{-1} \\ &= \frac{R}{1+j\omega RC} \\ &= \frac{R}{1+(\omega RC)^2} - j \frac{\omega R^2 C}{1+(\omega RC)^2} \end{aligned} \quad (4)$$

where,  $Z_R$  and  $Z_C$  are the resistive and reactive components.

This result directly indicates the form of real ( $Z_{re}$ ) and imaginary ( $Z_{im}$ ) impedance of the parallel RC circuit as:

$$\begin{aligned} Z_{re} &= \frac{R}{1+(\omega RC)^2} \\ Z_{im} &= \frac{C}{1+(\omega RC)^2} \omega R^2 \end{aligned} \quad (5)$$

and the phase angle  $\varphi$  can be presented as:

$$\varphi = -\omega RC \quad (6)$$

At low frequency ( $\omega RC \ll 1$ ),  $Z_{re} \approx R$  and  $Z_{im} \approx 0$ . This result implies that this RC circuit acts as a resistor. On the other hand, at high frequency ( $\omega RC \gg 1$ ),  $Z_{re} \approx 0$  and  $Z_{im} \approx 1/\omega C$ , the present circuit acts as a capacitor with time constant, equal to RC.

Eq. (5) and (6) yields

$$\left( Z_{re} - \frac{R}{2} \right)^2 + Z_{im}^2 = \left( \frac{R}{2} \right)^2 \quad (7)$$

Eq. (7) indicates a half-circle in the complex plane, with a radius of  $R/2$ , which can be validated by figure 6. The complex impedance plot also justifies the temperature dependency of AC conductivity and exhibits the same nature as given in figure 3.

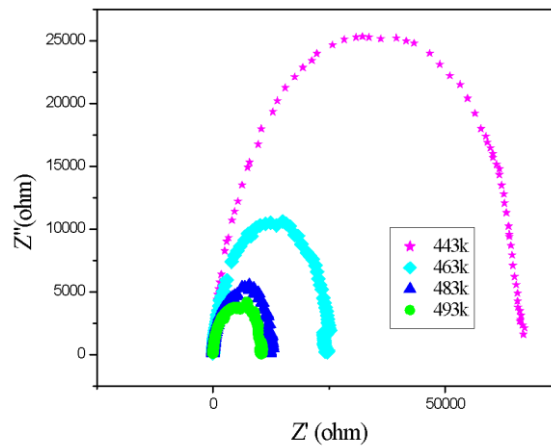


Figure 6: Complex impedance plot

#### IV. CONCLUSION:

Ionic conductivity of new glass ceramic containing  $\text{Li}^+$  has been studied. Temperature and frequency dependent conductivity have been interpreted using Almond-West formalism and Jonscher's universal

power-law respectively. The thermally activated nature of the sample is established by its study of DC conductivity ( $\sigma_{DC}$ ) and hopping frequency. The sub diffusive motion of  $Li^+$  ions is the cause behind the frequency independent conductivity (plateau region) in low-frequency zones. Whereas, the conductivity in high-frequency dispersive regions is because of correlated and pseudo-three-dimensional motion of  $Li^+$  ions in percolating networks. At lower temperatures, the AC conductivity is spreading out due to the dominant nature of AC conductivity over the DC conductivity. As the temperature is increased, the AC conductivity approaches the DC conductivity.

## REFERENCES:

- [1] B. Karmakar, K. Rademann, A. Stepanov Glass Nanocomposites: “Synthesis, Properties and Applications”, Elsevier 2016.
- [2] K. Takada, “Progress in solid electrolytes toward realizing solid state lithium batteries”, J. Power Sources, vol 394, pp. 74-85, August 2018.
- [3] W. B. Nasr, A. Mahmoud, F. Boschini and A. B. Rhaïem, “Optical and AC conductivity studies on  $Li_2-xRbx MoO_4$  ( $x=0, 0.5, 1$ ) compounds”, J. Alloys and Compounds, vol 788-803, pp.522-532, June 2019.
- [4] H. Pan, S. Zhang, J. Chen, M. Gao, Y. Liu, T. Zhu, Y. Jiang, “Li- and Mn-rich layered oxide cathode materials for lithium-ion batteries: a review from fundamentals to research progress and applications”, Mol. Syst. Des. Eng., vol 3, pp. 748-803, 2018.
- [5] X. Yao, B. Huang, J. Yin, G. Peng, Z. Huang, C. Gao, D. Liu and X. Xu, “All-solid-state lithium batteries with inorganic solid electrolytes: Review of fundamental science”, ‘Chin., Phys. B vol 25, pp.018802 ,2016.
- [6] E. R. Shaaban, M. A. Kaid and M. G. S. Ali,” X-ray analysis and optical properties of nickel oxide thin films”, Journal of Alloys and Compounds, vol 613, pp. 324-329, 2014.
- [7] A. Palui and A. Ghosh,” Structure-transport correlation of super-ionic mixed network former glasses”. Solid State Ionics, vol 343, pp. 115126 ,2019
- [8] S. Bhattacharya and A. Ghosh, “Relaxation Dynamics in Superionic Glass Nanocomposites”, J. Am. Cer Soc., vol 91, pp. 753-759, March 2008.
- [9] D. P. Almond, G. K. Duncan and A. R. West, “The determination of hopping rates and carrier concentrations in ionic conductors by a new analysis of ac conductivity”, Solid State Ionics, vol 8, pp. 159-164, April 1983.
- [10] D. L. Sidebottom, “New Materials for Optical Amplifiers” Phys. Rev. Lett., vol 83, pp.983-985, 1999.
- [11] S Bhattacharya, A K Bar, D Roy, M P F Graca and M A Valente, “Fractal dimensionality of ion conduction in glass nanocomposites” Mat. Phys. Mech., vol. 10, pp.56-61, August 2010.
- [12] X. Z. R. Yuan, C. Song, H. Wang and J. Jhang, Electrochemical Impedance Spectroscopy in PEM Fuel Cell: Fundamental and Applications, 2010, XII, 420 p, ISBN: 978-1-84882-845-2



**Aditi Sengupta** (is currently working as Assistant Professor of Electronics and Communication Engineering at Siliguri Institute of Technology and also is involved in Composite Material Research Laboratory funded by CSIR for her research-oriented works. The present research interests of Ms. Sengupta are Solid state Devices and Material Science, especially Nano Composites. Previously some research works on Image Processing has been done by her. The author is having publications in different National and International Conferences and Journals. She is passionate about Teaching and learning with total teaching Experience of 10 years.



**Sanjib Bhattacharyya** is recently working at North Bengal University as an Associate Professor of UGC-HRDC Department. Previously he worked as Assistant Professor at Siliguri Institute of Technology and Institute of Engineering and Management. He is also Principal Investigator of various major research projects, funded by CSIR and DST, Govt. of India. He published 50 papers in reputed international journal, Book and Book chapters and conference proceedings. 3 students have already completed their PhD works under his guidance and many more are working under his supervision. His keen interest is to develop new generation solid state materials for modern battery application, Glass and Glass-nanocomposites and Chalcogenide Glassy alloys.



Dr. Chandan Kumar Ghosh presently holds the post of Professor and HoD (R&D) in Dr.B.C.Roy Engineering College, Durgapur, India. He was in the post of HoD from 2012-2014 in the Dept. of ECE. He worked as Development Engineering in Sonodyne Electronics Co. Pvt. Ltd. and from 1996 to 1999 and worked as Assistant Manager (R&D) in Sur Iron & Steel Co.Pvt.Ltd. From 2000 to 2009 he was associated in the department of Electronics & communication Engineering of Murshidabad College of Engineering &Technology and from and. He published more than 70 contributory papers in referred journals and international conference proceedings. His current research interests include wireless communication, Array antenna using microstrip structures, MIMO antenna, DGS, EBG and microstrip resonator's integrated antenna structures.





## Experimental Study on Viscosity of Soybean Oil and its dependence on temperature

Tapas Roy<sup>1</sup>, Kamalika Tiwari<sup>1</sup>, Sunil Choudhury<sup>1</sup>, Santigopal Pain<sup>2</sup>

<sup>1</sup> Dr. B.C Roy Engineering College, Durgapur, West Bengal India

<sup>2</sup> Haldia Institute of Technology, Haldia, West Bengal, India

**Abstract.** In this work, the effect of temperature on soybean oil has been studied. The viscosity of five different brands of soybean oils is experimentally determined as a function of temperature (20°C to 90°C). The Viscosity of oil is an important physical property in processing and quality control. The experimental study shows a decrease in the viscosity of the oil samples with an increase in temperature. A negative nonlinear relationship between viscosity and temperature is observed.

**Keywords:** Soybean oil, Red Wood viscosity, temperature, Viscosity

### 1 Introduction

Vegetable oils are referred to those edible oils which are derived from plant sources. This oil contains fat-soluble vitamins thus contributing important sensory attributes to food products. Vegetable oils are responsible for providing growth and energy to our body. These oils are basically composed of glycerides of the fatty acids and are termed triglycerides [1]. Such oils have a wide range of applications in industries other than the food industry such as in the production of paints, varnishes, and lubricants. Some vegetable oils such as castor oil, linseed oil, Tung oil are particularly used as industrial oils, whereas oils such as palm oil, soybean oil, sunflower oil, and coconut oil have both industrial and edible purposes. Vegetable oil plays an essential role in food products as they are the carrier for fat-soluble vitamins such as vitamin (A, D, E, and K) [1-2].

Viscosity is one of the most important rheological parameters that illustrate the fluid texture. Edible oils generally have very high viscosity. Various factors that affect the viscosity of oils are additive, catalyst, temperature, shear rate, time, molecular weight, moisture, pressure, concentration, etc. Viscosity changes with shear rate, temperature, pressure, moisture, and concentration [3]. Among them, temperature and shear rate influences the viscosity the most. The viscosity of soybean oil is reported very high as compared to petroleum oils. The presence of unsaturated molecules in soybean oil is the root cause of high viscosity [4-5].

Past reports show an effect of temperature on the absolute viscosities of vegetable oils. The viscosity of oils and fats decrease linearly with temperature. The viscosity of corn oil with a capillary viscometer decreased with increasing temperature from 25°C to 80°C [6-7].

The present work reports the effects of temperature on the viscosity of different brands of soybean oil. The behaviour of different brands of soybean oil has been studied in the range of 20°C to 90°C. The viscosities of these oils are measured as a function of temperatures.

## 2. Materials and Method

### 2.1.1 Sample Collection

In the present experimental study, refined soybeans oil was purchased from the local market in Durgapur, West Bengal. Five different brands of refined soybean oils were selected based upon their commercial demand and are shown in Fig.1. All the brands are commonly used as cooking oil in the household. A programmable water bath was used to ensure precise control of temperature during measurements. All the measurements were carried out in triplet



Fig.1: Samples of different brands of Soyabean Oil

### 2.1.2 Sample Preparation

The oil samples were filtered using a fine filter cloth to remove suspended particles. All oils were kept at room temperature. The beakers were cleaned with alcohol. The temperature-controlled water bath was used to ensure stable and precise control of oil temperature during the measurement. Fig. 2 Shows samples of filtered soyabean oil for the experimental study.



Fig.2: Samples for experimental study

## 2.2 Experimental Setup

The experiment was carried out on a Redwood Viscometer to analyse the viscosity of different oil samples at different temperatures. The setup has a vertical cylindrical oil cup with an orifice in the centre of its base. The orifice is closed by a metal ball. The cylindrical cup is surrounded by the water bath. The viscometer is used to determine the kinematic viscosity of the oil.

The water bath maintains constant temperature by means of a programmable water bath. Thus, the temperature of the oil is maintained using the water bath. The heating profile for the temperature controller was set up from 20°C to 90°C in increments of 10°C & holding each temperature for 5 min. The water bath is stirred to maintain uniform temperature and record the temperature of the oil. The density is labelled in oil packs. The measurements were performed in triplet and values were considered to calculate the kinematic viscosity.



Fig.3: Experimental setup of Redwood Viscometer

## 2.3. Measurement of procedure using Redwood viscometer

The samples of each brand were divided into triplet before measurement. Each sample of 50 ml was pour into cylindrical cup. The orifice at the base was closed using ball valve. The water bath was filled with cold water. A thermometer is also used for monitoring the temperature. The oil is allowed to pass through the orifice. It is then collected in a flask. The time (sec) is noted to collect 50 ml of oil to pass through the orifice. The time taken to pass 50 ml of oil through the orifice is called redwood seconds. The above procedure was repeated at different temperatures and time was noted.

## 2.4. The temperature dependence of absolute viscosity

The redwood viscometer is used to determine kinematic viscosity of oil. The kinematic viscosity was measured using the equation (1).

$$y = (A \times t) - (B_ / t) \quad (1)$$

Table 1 shows the constants of Redwood Viscometer. The constant values were used in calculations of kinematic viscosity. The two different constants depend on the time of filling the 50 ml liquid in the flask.

Table 1: Redwood viscosity constant

Time of flow(sec)	A	B
Up to 100	0.0026 1.72	0.0026 1.72
Above 100	0.00247 0.5	0.00247 0.5

### 3. Results and discussion

#### 3.1. Relation between Temperature and Redwood time

The refined soybean oils which were analysed were from popular commercial brands. Table 2 shows the relation between temperature and redwood time (sec). The observation table shows the time taken to fill the 50ml flask by each brand of soyabean oil.

Table 2: Relationship of Temperature (°C) with Redwood Time (sec)

Sl. No	Temperature (°C)	Sample 1 (sec)	Sample 2(sec)	Sample 3(sec)	Sample 4(sec)	Sample 5(sec)
1	20	152.2	142.6	196.74	158.08	184.495
2	30	112.01	101.04	127.47	113.31	116.175
3	40	83.5	76.86	111.49	93	89.26
4	50	75.47	71.01	84.58	65.83	75.56
5	60	57.76	56.91	53.58	52.33	54.75
6	70	45.54	44.37	43.53	44.26	45.82
7	80	34.83	35.3	36.77	34.69	36.75
8	90	30.86	30.21	29.12	26.19	31.56

It is observed that the time taken by each brand of soybean oil sample decreases with rise in temperature. It is noted that all of the plots behave in similar patterns. It is suggested all the oils behave as Newtonian fluids [8-9]. Thus, the shear stress decreases with increasing temperature. This was due to the higher thermal movement between the oil molecules thus, reducing the intermolecular forces, causing the flow among them easier and a decline in viscosity is observed [10].

#### 3.2. Relation between Temperature and Kinematic viscosity

The kinematic viscosity of each brand of soybean oil is calculated using the equation (1). Table 3(a-e) shows the triplet based on observations of the redwood viscometer. It is observed that the soybean oil of different brands is having lower viscosity at the higher temperatures.

Table 3a: Sample 1

Temperature (°C)	Viscosity(TR1)	Viscosity(TR2)	Viscosity(TR3)
20	161.68	206.82	110.54
30	114.96	109.1	111.99
40	83.37	79.53	83.63
50	74	70.87	76.95
60	57.87	57.65	61.72
70	52.58	47.76	43.32
80	34.48	35.18	33.37
90	31.38	30.34	33.25

Table 3b: Sample 2

TEMP(°C)	VISCOSITY(TR1)	VISCOSITY(TR2)	VISCOSITY(TR3)
20	201.31	170.11	101.04
30	107.89	100.94	101.141
40	76.6	79.77	77.13
50	70.17	80.7	71.85
60	56.43	62.22	57.39
70	44.61	44.13	45.11
80	36.15	38.09	34.46
90	30.82	32.04	29.6

Table 3c: Sample 3

TEMP(°C)	VISCOSITY(TR1)	VISCOSITY(TR2)	VISCOSITY(TR3)
20	194.34	199.15	202.74
30	125.91	129.03	133.11
40	98.76	109.09	113.89
50	86.51	82.66	73.52
60	52.58	54.58	57.39
70	42.45	44.62	47.27
80	35.68	37.86	40.02
90	28.4	29.85	31.56

Table 3d: Sample 4

TEMP(°C)	VISCOSITY(TR1)	VISCOSITY(TR2)	VISCOSITY(TR3)
20	162.17	159.04	157.13
30	118.22	111.51	115.11
40	99.25	95.4	90.6
50	67.27	64.39	70.15
60	54.51	57.39	50.16
70	45.58	42.94	39.78
80	36.155	33.24	28.39
90	27.91	26.92	25.46

Table 3e: Sample 5

TEMP(°C)	VISCOSITY(TR1)	VISCOSITY(TR2)	VISCOSITY(TR3)
20	190.74	186.66	182.33
30	114.37	122.32	117.98
40	99	91.79	86.74
50	75.21	75.92	67.02
60	56.45	53.06	49.44
70	48.72	52.1	42.92
80	40.03	37.36	36.15
90	31.31	31.81	30.57

Fig.4 shows graph of the relationship between Kinetic viscosity of different brands of soybean oil at different temperatures. The plots indicate the viscosity tested for the oils were dependent on the temperature

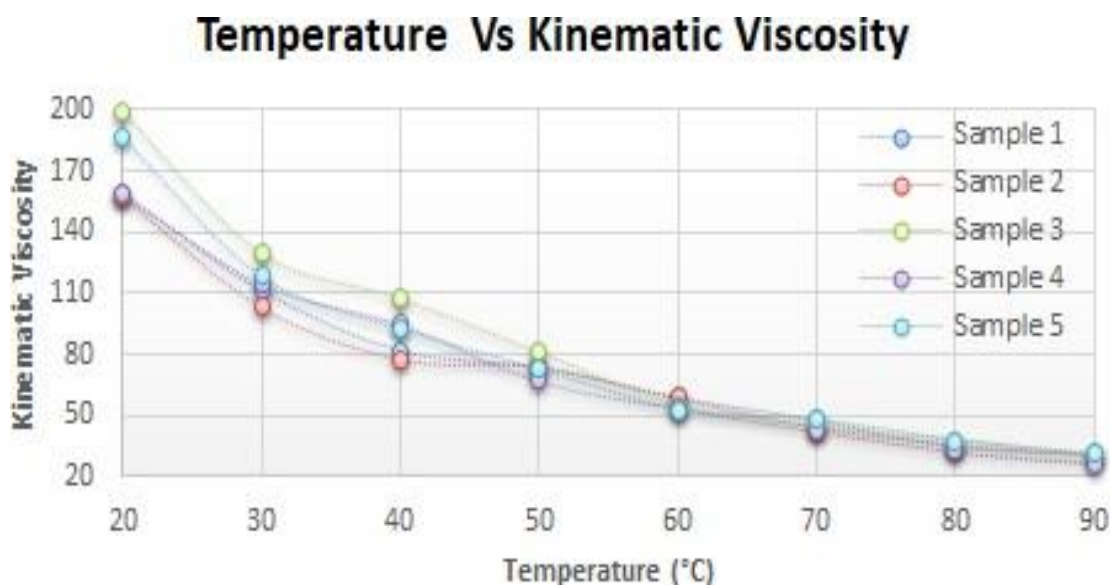


Fig.4: Graphical representation of Temperature Vs Kinematic viscosity

As the temperature of the oil decreases the viscosity increases. Also, the viscosity of the oil decreases with the elevated temperature of the oil under study. The viscosity values of all the brands of soybean oils decrease with increasing temperature. The experimental kinematic viscosities of different sets of oils were comparable with the published data even at different temperatures. The standard errors were very low, which indicates that the viscosity values obtained were very stable.

## 4. Conclusion

Viscosity is one of the vital rheological parameters for cooking oil. The experimental values of all the samples of soybean oils studied were found to be Newtonian fluids. Among all the samples, it is observed that the kinematic viscosity of the oils decreases as the temperature of the oil increases and vice versa. The rise in temperature causes a decrease in cohesive force within the molecules. The high frying temperature, the number of times of frying the fatty acid content of oils, decreases the oxidative stability and flavour quality of the oil. Also, the antioxidant decreases with the high frying temperature. Hence temperature of cooking oil and viscosity is very vital to be monitored.

Acknowledgments. The authors are thankful to Dr. B. C. Roy Engineering College for funding and infrastructural support of the experimental study.

## References

1. Igwe, I.O., "The effects of temperature on the viscosity of vegetable oils in solution," *Industrial Crops and Products*, vol. 19, pp. 185-190, 2004
2. Stanciu I., "A new viscosity-temperature relationship for vegetable oil" *Journal of Petroleum Technology and Alternative Fuels* Vol. 3, pp. 19- 23, 2012.
3. Khupase N., "Temperature Dependent Viscosity of Measures of Ionic Liquids at Different Temperatures", *Indian Journal of Chemistry*, Vol. 49, pp. 727-730, 2010.
4. O. O. Fasina and Z. Colley, "Viscosity and specific heat of vegetable oils as a function of temperature: 35° C to 180° C," *International Journal of Food Properties*, vol. 11, no. 4, pp. 738– 746, 2008.
5. Ioana. S., "Rheological Study Behaviour of Soyabean Oil," *Journal of Science and Arts*, vol. 3, No. 44, pp. 711-716, 2018
6. Oluwafunmilayo, A., Aworanti, Samuel E., Agarry , Ayobami O., Ajani., "A Laboratory Study of the Effect of Temperature on Densities and Viscosities of Binary and Ternary Blends of Soybean Oil, Soy Biodiesel and Petroleum Diesel Oil," *Advances in Chemical Engineering and Science*, Vol. 2, pp. 444-452, 2012
7. U. Schuchardt, R. Sercheli and R. M. Vargas, "Transesterification of Vegetable Oils: A Review," *Journal of Brazilian Chemical Society*, Vol. 9, No. 3, 1998, pp. 199- 210.
8. Ceriani, R., Goncalves, C.B., Rabelo, J., Caruso, M., Cunha, A.C.C., Cavaleri, F.W., Batista, E.A.C., Meirelles, A.J.A. *Journal of Chemical and Engineering Data*, 52(3), 965, 2007
9. Denton E. L., "Liquid Viscosity and Temperature", *California State Science Fair 2004 Project Summary*
10. Orthoefer, F.T., *Vegetable Oils*, in *Bailey's Industrial Oil and Fat Products*, 5th edn., edited by Y.H. Hui, John Wiley & Sons, New York, Vol. 1, 1996, pp. 19–44.



Tapas Roy: He received his B.Tech. degree from Kalyani University, West Bengal in 2004, M. Tech from Calcutta University (Raja Bazar Science College), Kolkata, West Bengal in 2006. He has been working as Assistant Professor in the department of E.C.E., Dr.B.C.Roy Engineering College since 2006 and having more than 15 years of teaching experience





**Dr. Kamalika Tiwari** : She was born on 31<sup>st</sup> March 1982 in Dhanbad, Jharkhand, India. She received her BE degree from Pune University, Maharashtra in 2005, PG Diploma in Steam Engineering from Forbes Marshall, Pune in 2006, M. Tech from NIT, Durgapur, West Bengal in 2009 and PhD from Jadavpur University, Kolkata, India. She was a Research Fellow at Jadavpur University and was associated with eAGRIEN funded project” Inward Leaf Inspection System for Indian Tea Industries” under Jadavpur University and CDAC, Kolkata. She received Young Scientist International Travel support under DST-SERB to South Korea in the year 2013. She had published several technical papers in SCI/Scopus journal and conference publications. Her area of research work is electronic tongue in agro based application, fabrication of electrochemical sensors based on Metal Oxide Nanoparticle, Molecular Imprinted Polymer and NIR spectroscopy in food quality assessment, Measurement and Instrumentation. She has been working as Assistant Professor in the department of AEIE, Dr.B.C.Roy Engineering College since 2006 and has more than 12.4 years of teaching and 2.4 years of research experience. She is Reviewer of Microchemical Journal, Elsevier and member of Technical committee of IEEE conference held at Dehradun in 2019 and reviewer of Advances in Sustainability Science and Technology, Springer 2021



**Sunil Kumar Choudhary**: He received his BE degree from Burdwan University, West Bengal in 2004, M. Tech. from NIT, Durgapur, West Bengal in 2007 and pursuing Ph.D. from NIT, Durgapur, West Bengal India. He is working as Assistant Professor in the department of E.E., Dr.B.C.Roy Engineering College since 2014 and having more than 14 years of teaching and 5 years of research experience



**Santigopal Pain** : He was born in Bankura, West Bengal, India in 1976. He received his B.E. degree from Jalpaiguri Govt. Engineering College (North Bengal University), M.E. degree from Jadavpur University and PhD degree from NIT Durgapur, India, in 1999, 2003 and 2018, respectively. He has two years industrial experience and 18 years of teaching experience. He was associated with BUIE, Bankura, and BCREC, Durgapur. He is currently working as an Associate Professor at Haldia Institute of Technology, Haldia, West Bengal. He holds several responsible positions like HOD, TIC of the institute, Dean etc. He has several National/International publications. He is the Editor of Journal of Advanced Research in Electrical Engineering and Technology and Journal of Advanced Research in Power Electronics & Power Systems, ADR Publication. He is also the technical committee member of several international conferences. He is the reviewer of IEEE System Journal, IET Generation, Transmission & Distribution Journal, ICCDC 2019 and ISET 2020. His research interest includes application of soft-computing techniques in Power System and Control System Engineering



# Hardware Trojan effect in Network on Chip: A brief view in security perspective

Banashree Chatterjee<sup>1</sup>, Monalisa Chakraborty<sup>2</sup>

<sup>1</sup>Department of Information Technology, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India,  
Email: [banashree2014@gmail.com](mailto:banashree2014@gmail.com)

<sup>2</sup>Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India,  
Email: [chakraborty.monalisa6@gmail.com](mailto:chakraborty.monalisa6@gmail.com)

**Abstract** —We have already entered modern technologies which have changed our life such as the internet, telecommunication, machine learning, AI. For developing a better system, we are increasing the number of transistors in that system, and this trend will be continued. Still, the main drawback of the weir connection is that it was quickly breakable that not reliable. Network-on-Chip Architecture was introduced to solve this problem and improve over conventional bus and crossbar communication architectures. After this, considerable modification of hardware during the fabrication process is called HT. This paper mainly focused on the area to detect and correct up to the two-bit error. Previously the hardware Trojan Detection on the NoCs is for the lesser message bit (i.e., 4-bit message bit and 4-bit parity bit).

It was capable of correcting the one-bit error and detecting the two-bit error. Hamming code (7, 4) is used. For error correction and detection, the extended Hamming code was used, which can correct the error up to one bit only. Shuffling and Reshuffling circuit was used for 8-bit transmission. The message bit was encoded using the Hamming code, which gets the flit from the Go-Back- N-Buffer. Here it is extended to detect and correct up to 2-bit errors using BCH code, and this can be extended up to three-bit mistakes. This code is used in telecom, wireless transmission, satellite. BCH (15, 7) is used to detect and correct the error flits. The code was written in VHDL on Xilinx 14.7 platform.

**Keywords** —*BCH, Hamming code, NoC, Shuffling and reshuffling Circuit.*

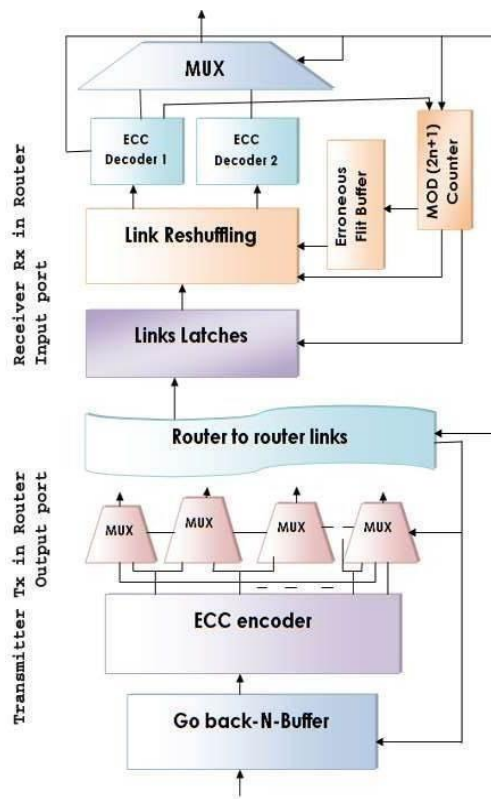
## 1. INTRODUCTION

In this age, 1990s, more processor spots and beast reusable components were amalgamated on silicon chips called structure on chip (SoC)[1][2][3]. The central troubles of this were interfacing virtually two plans and endorsing the entire structure concerning physical and veridical lead. In light of everything, they can relate possibly shy of 10 sodality limits and the conjecture of one structure concerning another cognate development was exasperating as the other correspondence partner utilized a proximate to kineticism. Starting there, the term Network on chip (NoC) is today used in incontrovertibly expansive level joining the hardware correspondence structure, the middleware, and working development sodalities, and plan framework and instrument to arrange onto an NoC[4][5][6]. A couple of indisputable processor living spaces, recollection focus, and application-unequivocal focuses are worked with miscommunication networks that are incidentally a task of NoC. In the encryption plot philosophy, when a bulwarked centre is relied upon to communicate with other centre information then it coalesced first. After that it was sent in an unmitigated given sodality. Predicated on security covering in which information was coalesced, unscrambled and message endorsement code (MAC) was set up by security

covering are advantages to increment the security of NoC Gebotys. For ignoring the DoS Trojan model affluence of work turns up at ground zero. For the space of Low force Trojan, Basim Shanyour Proposed standard cell situation procedure is utilized and it is shown low area overhead. Predicated on transfer speed comprehension information pack support plot was withal introduced. Each master's outcome showed 36% for fending move speed and 56% for low-area overhead. For the space and repulsiveness of sneaking around information, each analyst gave a strategy, and there they utilized the sneaking around invalidator module. In their arduousness results showed a 48.4% expansion in execution[7][8][9][10].

## 2. METHODOLOGY

We propose to detect remaining HTs at runtime using error control coding (ECC) and assume that noticeable HTs are filtered out by offline tester. To manage transient errors caused by particle strikes, crosstalk, and spurious voltage fluctuation in NoCs done by ECC. Whenever the trigger condition disappears Hts have become inactive. As well as the transient errors, HTs induce error also rare and volatile.



**Figure- 1: Proposed Transmitter (TX) router, & Receiver (RX)**

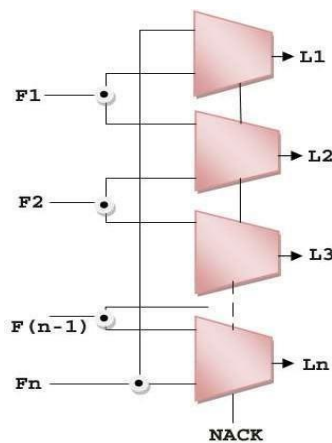
To add redundancy As on each uncoded flit we use an ECC Encoder which is clearly shown in Fig. 1. If the uncoded flit is new then it is from NoC arbiter otherwise the await flit came from Go-Back –N buffer. The syndrome and links are potentially controlled by HTs that are calculated with the help of an ECC decoder which is present in the receiver (Rx) router. There is some difference between transmitter router and receiver router like two ECC decoders are present in receiver router. In contrast, one encoder and one decoder are present in the transmitter router. To obtain an error-free flit in the link Reshuffling unit, those two ECC decoders cooperate.

Those flits affected by HTs are saved to EFB (Erroneous Flit Buffer) and triggered Counter MOD-(2N+ 1). For erroneous flit buffer, the counter generates a read signal after (2N+ 1) cycles and it also enables the link-reshuffling unit. The idle state occurs in the case of EFB and the mod-(2N+ 1) counter when an error-free flit is obtained.

**Link Shuffling Technique**

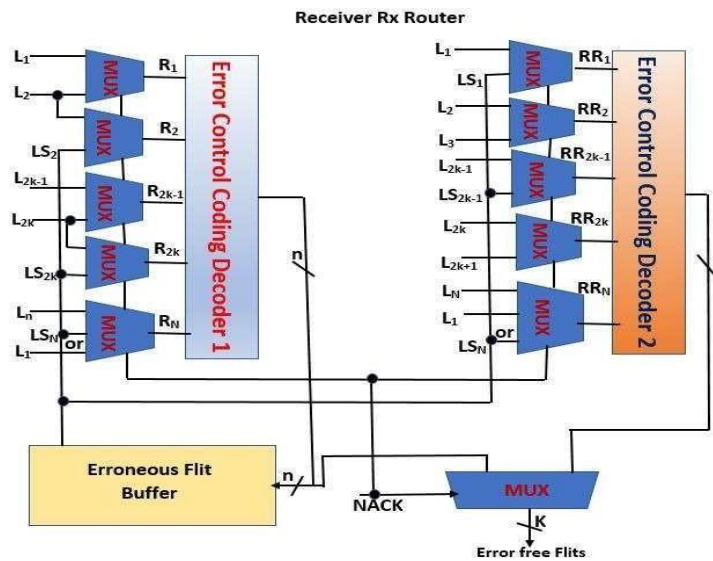
In the retransmitted flit avoid the HT affecting we need to switch the odd-even bits. No link shuffling will be required if it received transmitted flit is successful. If a non-zero syndrome obtains by the receiver router then the receiver will send a not-acknowledgment (NACK) to raise flit retransmission. Further retransmission required we can switch by the adjacent even and odd bits from the retransmitted flit. In Fig .2the odd bitF1, and even bit F2 .even link weir is L2 and odd link weir is L3. And the odd bit is connected to the even link wire (L2), and the odd link wire (L3) helps to transfer the even bit (F2). The main exception is that here the last bit Fm multiplexing with the first bit F1. After Shuffling the link, it will locate the failure link which is affected by the Trojan viruses. The 'NACK' signal from the router will make assure that the data bit transmitted is correct till that reshuffling of the bit will happen from 'even to odd' or vice versa.

The circuit diagram of the Shuffling Circuit is below:



**Figure-2: Shuffling Circuit**

### Reshuffling Circuit

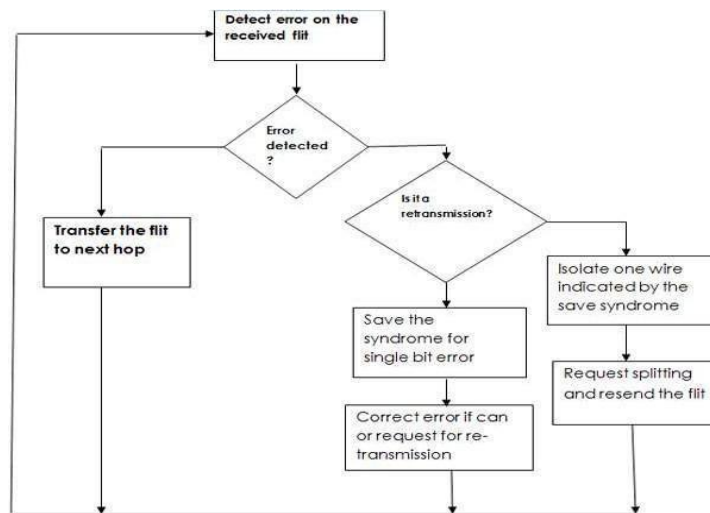


**Figure-3: Reshuffling Circuit**

To perform correction decoding and error detection on the receiver side, with previously saved flits the reshuffled flits are combined. This is clearly shown in Fig 3. After the received flit is retransmitted, then two flits are decoded simultaneously with the help of ECC decoders. The decode two flits are R and RR. Retransmitted flit can construct R flit with odd bit and it can construct the previous flit LS also. From the retransmitted flit The RR flit is produced with even bits and the RR flit is also produced odd bits from the previous flit LS. Fig.3 shows the detailed link reshuffling circuit. Both ECC decoder 1 and ECC decoder 2 inputs are reshuffled. The reshuffling has happened in several ways it helps the odd and even bit to appear perfectly aligned with the flit LS. The common case for ECC decoder 1 can be found for  $2k-1$  bits and the common case for ECC decoder 2 can be found for  $2k$  bits. At the time of using the original flits by the ECC decoder which is controlled with the help of NACK signal and the copy from the ECC decoder send to the next step. Activated HTs are addressed by the link reshuffling technique using NoC flit inputs. After the reshuffling procedure, the inputs of HTs are changed and during the period of retransmission inactivated by HTs scan. In the case of HT if it is on the even link wires, over the odd link wires the even bits of the flit will be retransmitted .like By putting Odd bit over even link wire, the HT controlling the odd link wires can be avoided by putting odd bits on the even link wires. The receiver router two copies of retransmitted flits are created. From those previous two copies, one copy would be selected for the next-hop transmission. The link reshuffling technique work together with a SECLUDED code can correct a single HT on link wires,

two HTs on odd link wires, and two HTs on even link wires.

The retransmission technique followed by reshuffling technique may not create a correct flit if one of the HT is on an odd wire and other HTs is on an even wire but the incorrect flit can be detected.



**Figure 4: Flow Chart of Link Isolation**

**Link Isolation Method**

The HTs are triggered by external signals that can remain for a long time. The short time HTs are effective by Link reshuffling which is followed with retransmission, but it is not appropriate for long-duration HTs the reason for that retransmission delay. To reduce the number of retransmissions we propose a wire isolation algorithm and it also helps to handle both HTs to separate as odd and even link wires.

**Link Isolation Algorithm**

Algorithm;

Step1. Code flit is input in the error detection transient method.

Step2. If the error is detected, either it will go for re-transmission or not

- a) if yes isolate the one wire indicated by the saved syndrome; Then request splitting transmission and resend the flit; Go to Step-1
- b) if No save the syndrome for single-bit error, correct the error if can or request for retransmission; Go to step1

Step3. if the error is not detected transfer the flit to the next hop.

### 3. RESULT ANALYSIS

#### Design of Encoder for (15, 7, 2) BCH Code

Encoder for (15, 7, 2) double error correcting BCH code is designed by organizing LFSR with generated polynomial  $1+x^4+x^6+x^7+x^8$  and implemented on Spartan 3AN XC3S700AN FPGA of Xilinx. The RTL view and Schematic are generated by synthesis with Xilinx ISE 14.7[10][11][12].

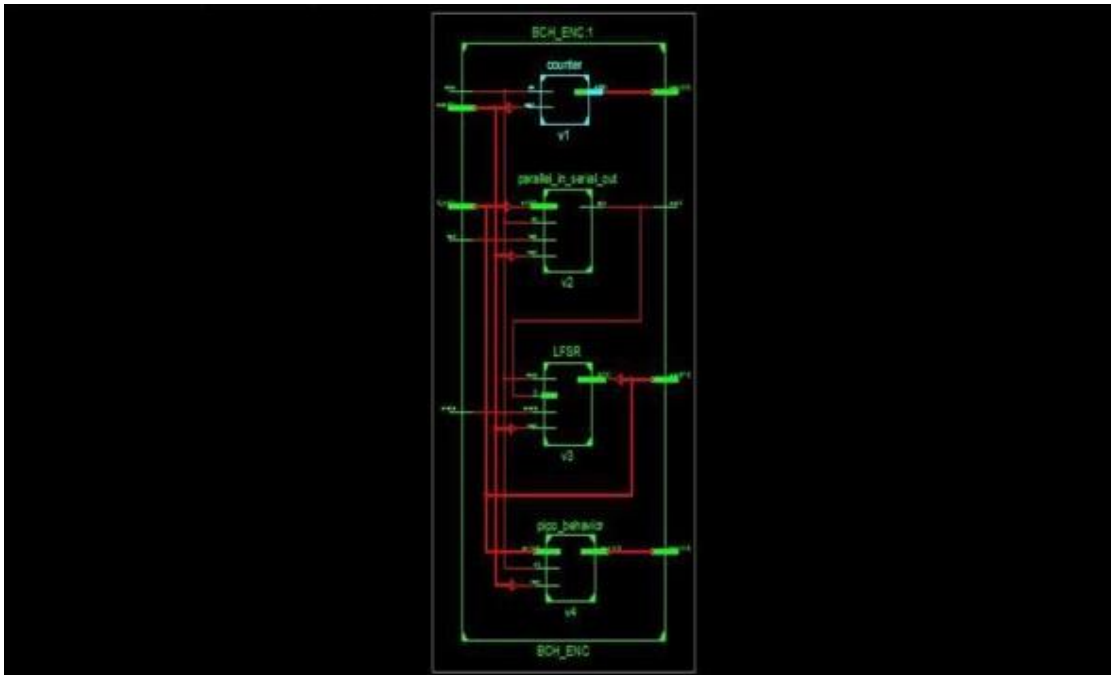


Figure-4 RTL Schematic of BCH ENCODER

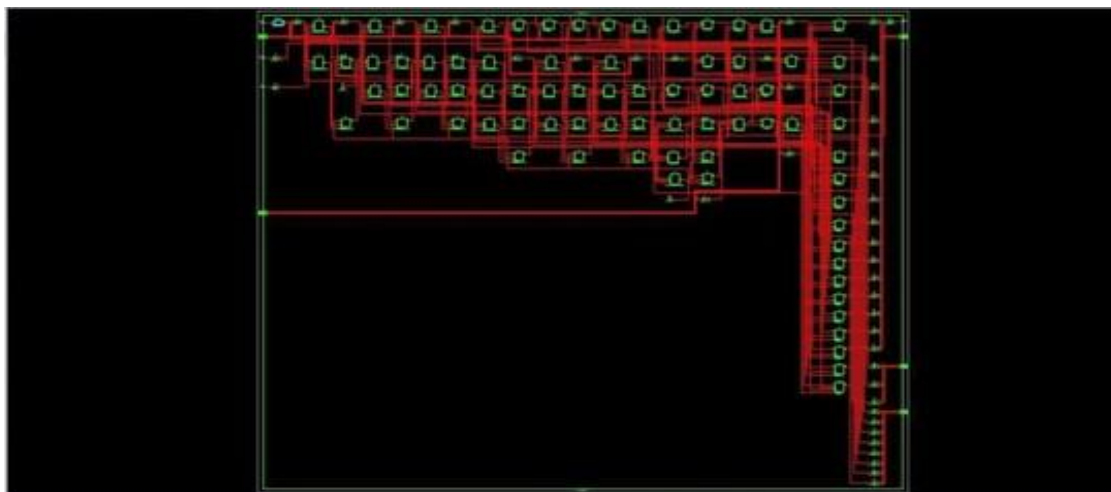


Figure-5 Technology Schematic of BCH ENCODER

#### Simulation Waveform Result of (15, 7, 2) BCH Encoder

The timing simulation of (15, 7, 2) BCH encoder is shown in Figure-6. A total of 9 clock cycles is taken to complete transmitting of 15 code words, 7-bits are information bit and 8- bits are parity bit. 7 Information

bits "1101101" are transmitting as it is whereas other 8 bits "10110110" are transmitting. The Value Marked in the yellow circle is the message bit, whereas the value marked in red is the Encoded Message bit.

So Encoder output bits are "110110110110110"



Figure-6 Simulation Result of BCH ENCODER

### RTL Schematic of Shuffling Circuit:

Shuffling Circuit is used to shuffle the link to detect which link is maliciously affected by Trojan viruses. In this design, fifteen number of 2:1 multiplexer and one number of two input Or-gate is used. VHDL code for multiplexer and or-gate is written in Xilinx-14.7. All the components are called in the main program to design the Shuffling Circuit.

The RTL schematic of the shuffling Circuit is shown below figure-7. This design is used after the Encoder design, the encoded message from the encoder is fed into this design. The combination of the encoder and the shuffling Circuit is in the transmitter side of the Hardware Trojan Detection System.

### Simulation Result of shuffling Circuit:

Input from the encoder to shuffling Circuit is "110110110110110".

When the NACK signal from the router is logical "Zero" or '0'. It will shuffle the message bit otherwise it will directly transmit the encoded message to the router. In the figure-8 the circled value in yellow is encoded message bit input to the shuffling Circuit and the value circled inside red is output from the shuffling Circuit.



Hence the output is 101101101101101 when NACK is logical "zero" or '0'. We can see that the message bit is shuffled.



Figure-7 Schematic of Shuffling circuit

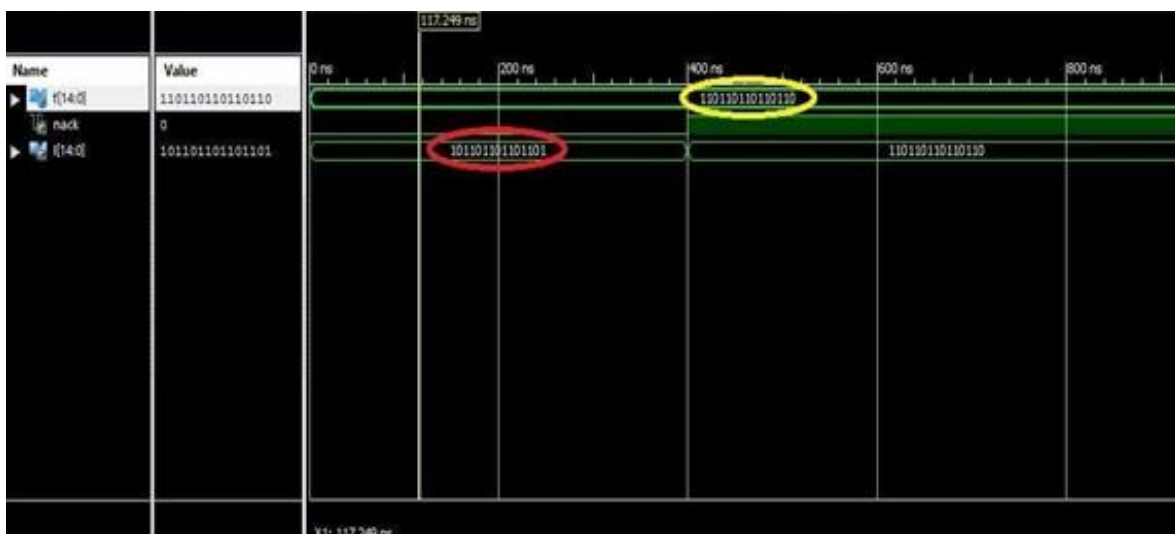


Figure-8 Simulation Result of Shuffling Circuit

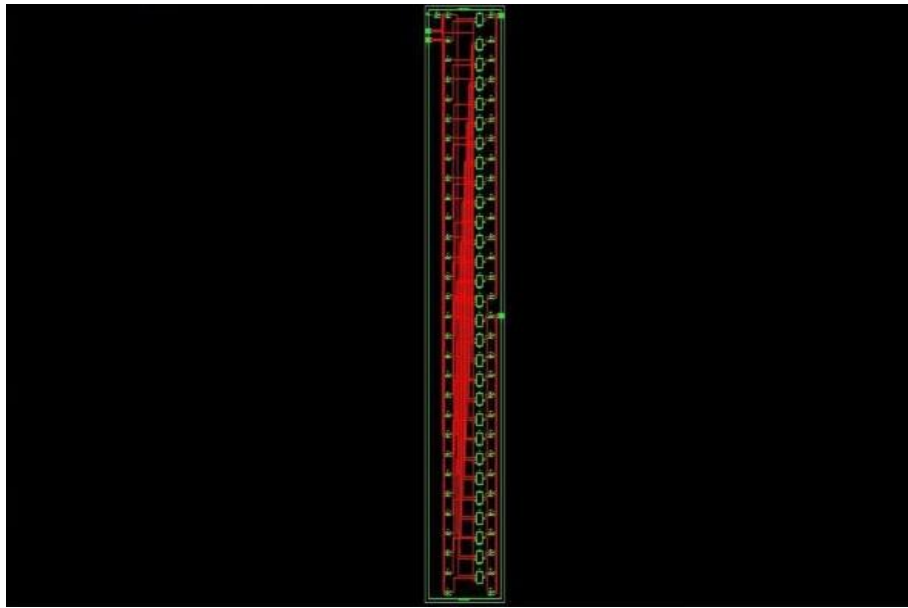
Input from the encoder to shuffling Circuit is "110110110110110". Output from the shuffling Circuit is "110110110110110". Input is shown in value circled by yellow whereas the output is by red in the figure-9. Hence the message bit is shuffled when NACK signal is "zero" otherwise not shuffled.

**RTL Schematic of Reshuffling Circuit:**

This Reshuffling circuit is used to reshuffle the received message bit if there is some link that is affected by Trojan viruses otherwise it will simply receive the message bit will send it to the decoder to check if the reins some error. There shuffle Circuit will reshuffle the message bit until it received the error-free flits. For this design, we have designed it for 15-bit data. Hence, we needed 30-multiplexer and two or-gate. The components above mentioned are written in VHDL on the Xilinx-14.7 platform and the components are called in the main program to design the Reshuffling circuit. It is basically on the principle of odd and even bit shuffling. One shuffling Circuit will shuffle one even bit other on an odd bit. So, the output from the Even bit shuffling is sent to the decoder-1 and the output from the debit side coder-2.CodewritteninXilinx-14.7andRTLschematic of reshuffling circuit is generated. In the figure-10 the RTL schematic of the reshuffling circuit is shown.



**Figure-9 Simulation Result of Shuffling circuit**



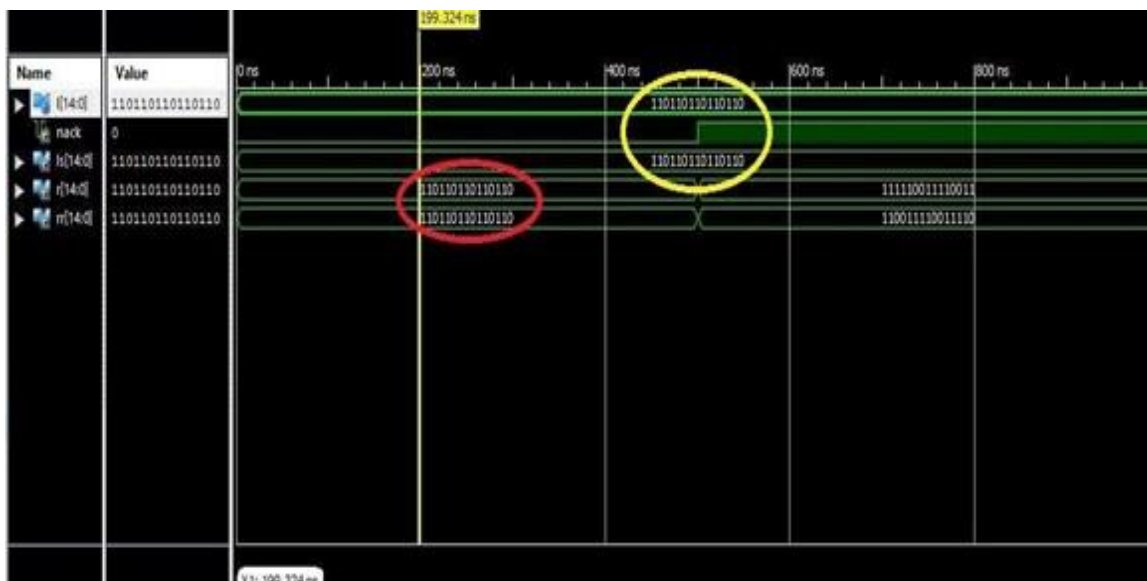
**Figure-10 Schematic of Reshuffling circuit**

**Simulation Result of reshuffling circuit:**

In the below simulation result the input is shown in the value circled inside the yellow and the output is shown in the red circled value. We can see that when the NACK signal is logical "zero" or '0' the output is the same as input since no reshuffling is done as the message bit is error-free received. The simulation result of the reshuffling circuit is shown in figure-10.

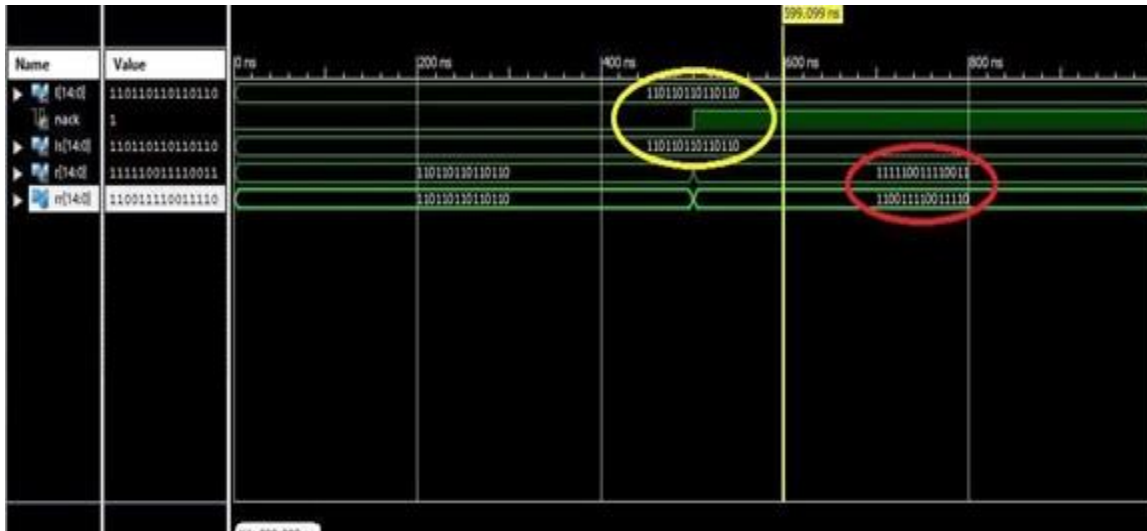
**Simulation Result of reshuffling circuit:**

In the below simulation result the input is shown in the value circled inside the yellow and the output is shown in the red circled value. We can see that when the NACK signal is logical "zero" or '0' the output is the same as input since no reshuffling is done as the message bit is error-free received. The simulation result of the reshuffling circuit is shown in figure-11



**Figure-11 Simulation Result of Reshuffling circuit**

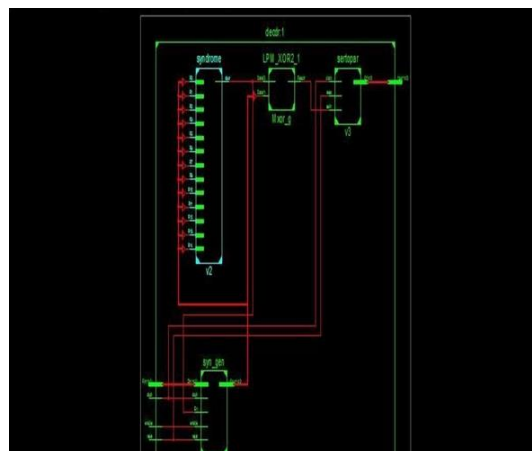
In the below simulation result the input is shown in the value circled inside the yellow and the output is shown in the red circled value. We can see that when the NACK signal is logical "one" or '1' the output is shuffled since reshuffling is done as the message bit is erroneously received. The simulation result of the reshuffling circuit is shown in figure- 11



**Figure-12 Simulation Result of Reshuffling circuit**

### RTL Schematic of BCH Decoder (15, 7) Code

The decoder for (15, 7, 2) double error correcting BCH code is designed by organizing Parallel to serial Shift Register, syndrome generator, error locator, and error correction using Majority gate. All the components are called in the main program to design the BCH decoder and implemented on Spartan 3AN XC3S700AN FPGA of Xilinx. The RTL Schematic is generated by synthesis with Xilinx ISE14.7.



**Figure-13 Schematic of BCH DECODER circuit**

### Simulation Result of Decoder BCH (15,7)Code

The timing simulation of (15, 7, 2) BCH Decoder is shown in Figure. A total of 15 clock cycles is taken to complete transmitting of 15 code word, 15-bit received code word bit . At 1500ns, we can see the corrected data. The Value Marked in the yellow circle is the message bit, whereas the value marked in red is the corrected bit. In this simulation, we have intentionally input the error to check it is correct or not. So, from the extreme left bit at the 3<sup>rd</sup> and 6<sup>th</sup> position, we have changed the bit value. Since Encoder output bits are "110110110110110" We input to the Decoder "111111110110110". We get the corrected code, marked in the red circle.

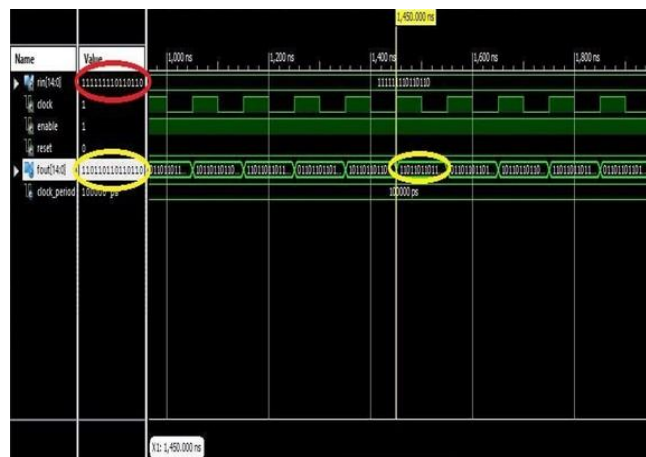


Figure-14 Simulation Result of BCH DECODER

### 4. CONCLUSION

Hardware attack of malignant Trojans (HTs) is preposterous chip security. As NoCs become a famous on-chip interconnect structure for multi-/many-focus systems, HT conspicuous support for NoCs is pivotal. To encompass the ideal chance for dissevered HT testing, we foster a bumble control procedure to address HT-commenced blunders at runtime. Goof control coding is abused to apperceive the HT screws up at the information and yield ports of NoC switches. To address HTs' supersession and conflictingly flighty accentuation, we recommend an early sodality upgrading procedure to profit from the abundance alive skip retransmission plan. In this Final Thesis report, the re-endeavouring and reshuffling circuit utilizing the Xilinx-14.7 stage is composed. Sundry areas to optically outwardly inspect how transmission interface is spread on the off chance that it is Hardware Trojan . Sundry potential areas yield is depleted eyewitness to utilizing VHDL coding. Each part utilized in the upgrading Circuit is checked. For the 15-piece upgrading and reshuffling circuit, 45 Multiplexer, and two or-ingresses are utilized. Stark mix-upchange evaluations and plans are habituated to apperceive and outline the single-piece goof and agonize the twofold piece screw up are stood disunited from BCH code capable with a transmutation up to 2-mess up. This report is prevalently fixated on the BCH code computation. Reshuffling system is utilized plerarily for brief length Hardware Trojans (HTs). Like this, interface disunion estimation is supplementally used to initiate the results of long-range HTs. In future work, we can execute better botch inspecting code that can disentangle more than no-account mishandles, so retransmission of message bit isevaded.

### References

[1] M. Tehranipoor and C. Wang, "Introduction to hardware security and trust," *Introd. to Hardw. Secur. Trust*, vol. 9781441980809,

- pp. 1–427, 2012, doi: 10.1007/978-1-4419-8080-9.
- [2] S. Narasimhan, W. Yueh, X. Wang, S. Mukhopadhyay, and S. Bhunia, "Improving IC security against trojan attacks through integration of security monitors," *IEEE Des. Test Comput.*, vol. 29, no. 5, pp. 37–46, 2012, doi: 10.1109/MDT.2012.2210183.
- [3] I. J. M. Wesseldijk-Elferink, A. W. Hendriks, and S. C. G. H. van den Heuvel, "Shared decision making in a semi-secluded chronic psychiatric ward: The reflective lifeworld experiences of patients with schizophrenia or schizoaffective disorders and nursing staff," *Arch. Psychiatr. Nurs.*, vol. 35, no. 5, pp. 519–525, 2021, doi: 10.1016/j.apnu.2021.07.007.
- [4] A. Y. Romanov, "Development of routing algorithms in networks-on-chip based on ring circulant topologies," *Heliyon*, vol. 5, no. 4, p. e01516, 2019, doi: 10.1016/j.heliyon.2019.e01516.
- [5] A. Y. Romanov, E. V. Lezhnev, A. Y. Glukhikh, and A. A. Amerikanov, "Development of routing algorithms in networks-on-chip based on two-dimensional optimal circulant topologies," *Heliyon*, vol. 6, no. 1, p. e03183, 2020, doi: 10.1016/j.heliyon.2020.e03183.
- [6] S. Bhattacharya, S. Bhattacharjee, A. Das, A. Mitra, and I. Bhattacharya, "Machine learning-based Naive Bayes approach for divulgence of Spam Comment in Youtube station," *Int. J. Eng. Appl. Phys.*, vol. 1, no. 3, pp. 278–284, 2021.
- [7] H. Vincent, L. Wells, P. Tarazaga, and J. Camelio, "Trojan Detection and Side-channel Analyses for Cyber-security in Cyber-physical Manufacturing Systems," *Procedia Manuf.*, vol. 1, pp. 77–85, 2015, doi: 10.1016/j.promfg.2015.09.065.
- [8] F. Nejadmoghadam, A. Mahani, and Y. S. Kavian, "A New Testing Method for Hardware Trojan Detection," *Procedia Technol.*, vol. 17, pp. 713–719, 2014, doi: 10.1016/j.protcy.2014.10.197.
- [9] H. A. M. Amin, Y. Alkabani, and G. M. I. Selim, "System-level protection and hardware Trojan detection using weighted voting," *J. Adv. Res.*, vol. 5, no. 4, pp. 499–505, 2014, doi: 10.1016/j.jare.2013.11.008.
- [10] S. Bhattacharyya, A. Misra, and K. K. Sarma, "A BCH code assisted modified NCO based LSPF-DPLL topology for Nakagami-m, Rayleigh and Rician fading channels," *Digit. Commun. Networks*, vol. 5, no. 2, pp. 102–110, 2019, doi: 10.1016/j.dcan.2017.10.001.
- [11] H. Liu, C. Ding, and C. Li, "Dimensions of three types of BCH codes over GF(q)," *Discrete Math.*, vol. 340, no. 8, pp. 1910–1927, 2017, doi: 10.1016/j.disc.2017.04.001.
- [12] J. Restrepo, E. Zemanate, N. Rodig, G. Daouk, and J. Herrin, "Pos-682 Sister Transplant Center: Bch-Usa and Fundacion Valle Del Lili-Fvl Colombia C To B Level 2016-2019," *Kidney Int. Reports*, vol. 6, no. 4, p. S298, 2021, doi: 10.1016/j.ekir.2021.03.713.



**Banashree Chatterjee** (Assistant Prof, IT department, Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)



**Monalisa Chakraborty** (Assistant Prof, CSE Department, Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)



## Machine Learning-based Support Vector Machine (SVM) approach for spam requests is tangible proof of computerized security

Priyanka Dhara<sup>1</sup>, Sohom Bhattacharya<sup>1</sup>, Shubham Bhattacharjee<sup>1</sup>, Subir Gupta<sup>1\*</sup>

<sup>1</sup>Department of Masters of Computer Application, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [priyankadhara052@gmail.com](mailto:priyankadhara052@gmail.com)

<sup>1</sup>Department of Masters of Computer Application, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [sohombhattacharya121@gmail.com](mailto:sohombhattacharya121@gmail.com)

<sup>1</sup>Department of Masters of Computer Application, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [shubhambhattacharjeebabai@gmail.com](mailto:shubhambhattacharjeebabai@gmail.com)

<sup>1</sup>Department of Masters of Computer Application, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [subir2276@gmail.com](mailto:subir2276@gmail.com)

**Abstract** — Younger generations are becoming more reliant on modern online media. It has some positive and negative effects. Various organizations are attempting to reach out to the masses. It is also possible to make significant progress at this stage. Considering all things, each innovation in science has some repulsive effect, and this terrifying effect gets additionally caused by human information. Facebook, Twitter, Instagram, YouTube, Email, and other commonly used automated stages are available today. Because it is filling the world indefinitely, usage is rapidly increasing, and the opposite way spam or ham is similarly extending it. As a matter of first importance, we must define spam as any unnecessary or unwanted message, such as a direct message or any pictorial message with an attachment, or any other type that has a dreadful effect in this high-level world. Spam can make correspondence a nuisance and, ultimately, a pointless activity that is often infected with malware that can damage our computers. The standard comment or message that we give or attach to any level of the stage is, of course, ham. There are a few ways to get the hang of programming in which we can see spam or ham at a prominent stage. Machine learning is probably the most adaptable investigation space in the current world. Because of its barrenness and worth, we attempt to perceive spam remarks utilizing this framework. In machine learning, there are a couple of computations. Nevertheless, we have taken the Support Vector Machine (SVM) since it performs well on the unquestionable dataset. This paper addresses the machine learning-based Support Vector Machine approach for separating spam and ham comments. 40 k datasets evaluate the presentation of the proposed procedure. The results show the prevalence of our technique by achieving a precision of 96%. The results in like manner reveal the amplexness of the proposed model similarly as accuracy score, precision score, recall score, F1 score, AUC score, model training time, and mean squared error.

**Keywords** — *Chi-Square Test, Ham Comment, Machine Learning, Spam Comment, Support Vector Machine*

\*Corresponding author: Subir Gupta, [subir2276@gmail.com](mailto:subir2276@gmail.com)

## 1. INTRODUCTION

Nowadays, websites play an essential role in the need for people to adequately express their views and share their evaluation of things and relationships through the web, business protests, conversations, and web diaries[1][2][3]. The vast majority read audits about items and administration before getting them. Everyone in the digital world is presently recognizing the significance of these online surveys for different clients and sellers. Merchants are likewise fit for planning their extra showcasing methodologies dependent on these audits. The maker may be cautious and address this problem to gather client devotion. Of late, the case of spam assault has expanded, considering how any end client may handily become a part of a spam audit and post it to the web business areas with no hindrance. Any organization may employ people to compose counterfeit audits for their items and administrations; such individuals are called spammers[4][5][6][7]. Spam audits are typically written to obtain a benefit or advance their products or services[8][9]. This training is known as survey spamming[10]. One of the significant issues with evaluating sharing locales is that spammers can unquestionably cause exposure to a specific thing by creating spam overviews. These spam overviews may play an indispensable part in extending the value of the organization. All the evidence suggests that spam surveys have become a significant issue in internet shopping, posing a risk to both the client and the maker. Audit spam can monetarily influence organizations and might cause a feeling of doubt in the overall population. Thus, because of its importance, this issue has pulled into the thoughts of the media and governments. Subsequently, recognizing spam surveys seems, by all accounts, to be a key region, and without settling this significant issue, online audit destinations could turn into a spot brimming with lies and futile. There have been a couple of research papers in assessment mining up to this point. The focus of these examinations has only been on Facebook posts, Twitter data, conversation groups, newsgroup posts, E-mail spam, and spam survey location models when all is said and done. In contrast, the current research likewise centers around openly accessible audit datasets and highlighting Notwithstanding that, there are still a ton of opportunities to get better at spam survey identification procedures. Finally, extraordinary audit spam discovery procedures, such as Lexicon-based methods and machine learning, are used to determine which surveys are spam[11][12].

Machine learning has found its way into practically every computational task on the cutting edge, particularly in the fields of information sifting and PC vision, where it is challenging to create calculations that are customary to carry out the required undertakings or work. Machine learning is perhaps the most sought exploration because of its expanded Degree and pertinence. Machine learning is likewise used to recognize spam remarks[13]. Furthermore, arbitrary wood calculations, naval force predisposition algorithms, data science, material science, engineering, Biomedical Engineering, and so on are available in this regard[14][15][16]. The Support Vector Machine is a predictive investigation used for spam comments in this study[17][18].

## 2. METHODOLOGY

Figure 1 portrays the strategy used to agonize spam remarks during the time spent doing them. For any test assessment, the dataset must manage a solitary eminent dataset. An assortment of approaches is required. The most crucial errand is to accumulate datasets from open-source sources. The data in the graph is collected from two open sources. One is GitHub, and the other is Kaggle's, and all the data is likewise handled physically. After



amassing the data, it is shipped off to the subsequent stage. Following the information stockpiling in the information store measures comes the data preprocessing stage. The accumulated dataset is exposed to three stages of the whole denominated data preprocessing, which are depicted as follows –

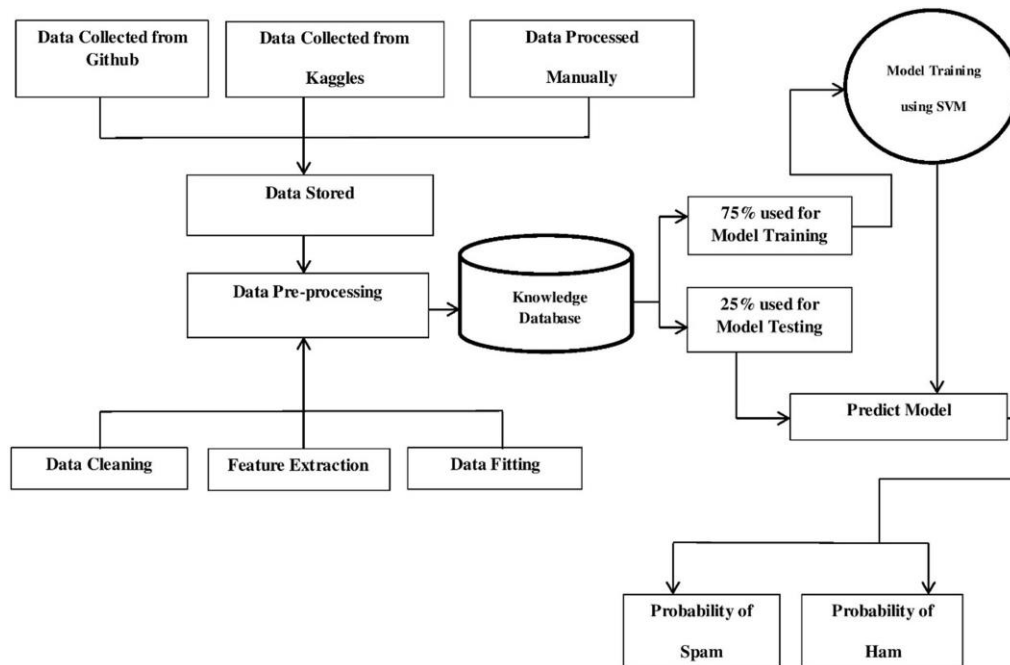


Figure 1- Spam Comment recognition in Machine Learning through Support Vector Machine classifier

The main piece of the interaction with the location of spam is the preprocessing of data or, in any case, it is likewise designated as data cleaning. This is when unstructured data is transformed into more organized information. This consumes the assistance in sorting through all of the data. We are only concerned with the remark section of the dataset in our model, so we have fixated on determining the presence of undesirable constructions in that segment. After the cycle of data cleaning, we have concocted clean data, which gets further utilized in the following cycles [19-21].

Following that is the highlighted extraction period, which significantly impacts the separation of spam and ham comments. In the Machine Learning dimension, the element extraction measure begins with an underlying arrangement of estimated information. It constructs highlights or inferred values that are proposed to be non-redundant, just as instructive in this manner, working toward getting the hang of driving to better expectations. When we input information into a Machine Learning calculation that is too enormous to even think about getting prepared, it gets suspected as excess data. It then, at that point, gets changed or decreased into a bunch of highlights that are named, including the vector. In this model, we have utilized the CountVectorizer, which is used to change a book into a vector dependent on the recurrence or check of every one of the words in the content. The CountVectorizer is used to create a lattice in which a segment of the lattice addresses each novel word, and each cell esteem addresses the word count comparison [22-23].

The interaction of data fitting comes into the job after including extraction. We need to develop the information into a particular construction before taking care of its anything but a classifier and building a model out of it. This progression requires the normalization of information which is the interaction of re-scaling the characteristics with the goal that they can have meant as 0 and change 1 to cut down the highlights to a typical scale however not mutilating the distinctions. The fit transform () that is utilized on the preparation dataset is being used for scaling it and finding out about the scaling boundaries of the dataset. Hence, the model we worked on will learn the mean and difference of each element of the preparing dataset. It includes two capacities – fit () for figuring the mean just as the fluctuation of the highlights present in the information and the change () for the shift in that load of highlights by the use of their individual mean and the corresponding variances [24-25].

The data created after data fitting is stored in the knowledge dataset. Along these lines, 40k notwithstanding datasets are held in this knowledge dataset. This dataset is split into two classes, where 75% of the data is supervised for model training and at some point, later another 25% of the dataset is overseen for testing purposes. The training dataset is then taken care of into Support Vector Machine Classifier for spam or ham remark arrangement.

Support Vector Machine is a kind of managed Machine Learning approach that can be utilized to address characterization and relapse issues. It utilizes the piece stunt to change the information and afterward makes an ideal limit between the potential yields dependent on these changes. In this proposed model the Support Vector Machine pseudo code in figure 2. For this training purpose, some assumption is considered:

1. Each data point can be represented as an n-dimensional vector.
2. The training data can be linearly separable or non-linearly separable.

Pseudocode for the SVM is given underneath:-

- Step1. At first, start extracting the features in the document and create a feature vector out of it.
- Step 2. Next step involves the fitting of those features in the form of the standard equation of a straight line.
- Step 3. Define a regularization parameter that provides an optimization function and a gamma parameter that defines the degree of influence that a training example can have in the classification problem domain.
- Step 4. There can be many equations that could represent the set of features in the n-dimensional space. Find the best hyperplane that will serve as the decision boundary for classifying data points.
- Step 5. Then find the vectors that are nearest to the chosen hyperplane and are critical in defining the position of that hyperplane. These are termed the support vectors.
- Step 6. Then attempt to read each sample input and extract the features available in them into a suitable form.
- Step 7. Prediction is made based on the pre-defined hyperplane and support vectors.
- Step 8. If the predicted outcome matches the actual output then, the model's parameters do not get adjusted.
- Step 9. If the predicted outcome does not match the actual output then, the model's parameters are updated to ultimately update the hyperplane and support vectors.
- Step 10. Steps 6 to 9 need to be executed iteratively unless the best hyperplane is achieved.

Figure 2- Pseudo Code for Support Vector Machine Classifier

In figure 2 from the commencement, we begin disconnecting the report's highlights and making a section vector out of it. The resulting stage consolidates the fitting of those highlights as the standard condition of a straight line. We require to depict a regularization limit that gives a redesign work and a gamma limit that represents the caliber of impact which an arranging model can have in the solicitation issue area. There can be different conditions that could address the arrangement of highlights in the n-dimensional space. We must track down the best hyperplane that will fill in beyond what many would consider possible for mentioning server farms. Then, at that point, we are required to discover the most proximate

vectors to the picked hyperplane and are central in depicting the condition of that hyperplane. These are designated as the avail vectors. Then, at that point, we endeavor to optically canvass each model's information and concentrate the highlights open in them into a plausible plan. Conjecture is made ward on the pre-portrayed hyperplane and support vectors. On the off chance that the mundane result orchestrates with the certifiable yield, as far as possible, don't get transmuted. Expecting the expected result doesn't sort out with the exact product, as far as possible are resuscitated to at long last potentiate the hyperplane and sponsorship vectors. Steps 6 to 9 should be executed iteratively except if the best hyperplane is refined.

After that, when the model is anticipated, the testing dataset is shipped off. This prognosticate model's precision is adscititiously determined near the cessation of this interaction. After agreeable outcomes have been engendered, we perpetuate onward to preserve the model to be utilized later on for adventitious upgrades. We use a model for this situation and convert it into anything but a byte stream and the other way around, according to the requisites of the problem. The last advance in this system is to utilize the preserved pickle model for testing prospects on examples given unequivocally from the outside. We opened the pickle spam grouping model and used it to anticipate the comments given to decipher what class the remarks belong to. For example, we performed a minuscule unit testing class to ascertain that the model we engendered has been developed felicitously and consistently, assures palatable outcomes.

### 3. RESULT ANALYSIS

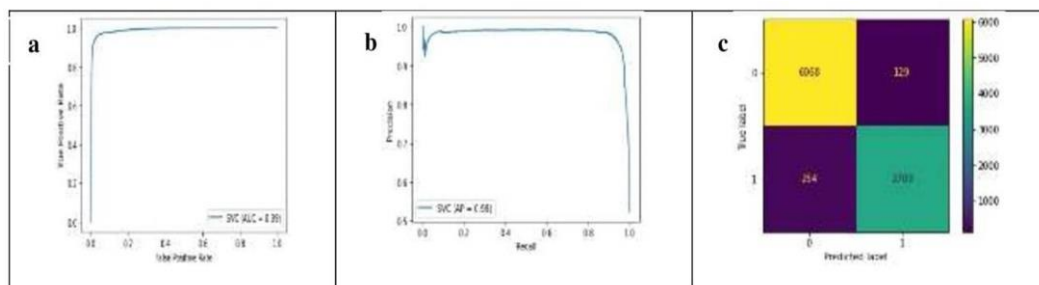


Figure 3- Measurement of (a) AUC Curve (b) Precision-Recall Curve (c) Heatmap/Accuracy graph of using Support Vector Machine Classifier

Figure 3(a) shows the AUC outfits a hard and expeditious degree of execution across all reasonable solicitation limits. One strategy for explicating AUC is the likelihood that the model positions a self-assured positive model more hugely than an aberrant negative model. Utilizing AUC work, we get 99.0255% (up to 4 decimal numbers) precision of the system shown in table 1. For precision checking, chi-square hypothetical testing and mean square error (MSE) were similarly decided. In MSE results emerged at 0.0377 (up to 4 decimal numbers)[26]. On the other hand, under Chi-square hypothetical testing considering notional theorization that is

Null Hypothesis H0: There is no connection among Content and Class relegations,

Alternative Hypothesis H1: There is a connection between Content and Class relegations

With considering of Level of paramountcy is 0.05, the Critical value is 35928.3509 (up to 4 decimal numbers), and the Degree of liberation is 35489, Chi\_square\_statistics found is

40639.0000 (up to 4 decimal numbers) shown in table 1, which infers Null hypothesis Reject and alternative theory acknowledged[27]. Figure 3(b) addresses a Precision-Recall Curve. Bountiful like the ROC curve, the precision review wind is applied to investigate equipollent solicitation calculations' presentation. It is customarily utilized in conditions where classes are emphatically transmuted. Adscitiously, as ROC turns, precision review turns to supply a graphical portrayal of a classifier's show across different abstract focuses instead of alone worth. Customarily in decided fall away from the faith, if understanding is relied upon to have a spot with the positive class at likelihood > 0.5, it's anything but a checked assurance. After a short time, we could pick any likelihood limit somewhere in the extent of 0 and 1. A precision review turn allows imagining what the edge decision denotes for classifier execution and can even avail us with picking quite far for a particular issue. In this methodology, we discover the F1 score is 95.0904% (up to 4 decimal numbers), shown in table 1.

Table 1- Outcome limit of the Support Vector Machine Classifier

Algorithm	Support Vector Machine (SVM)
Chi-Square Test	40,639.0000
Accuracy score	96.2303
Precision score	96.6389
Recall score	93.5907
F1 score	95.0904
AUC score	99.0255
Model training time	189.1747
Mean Square Error	0.0377
Critical value	35928.3509
Degree of freedom	35489

Figure 3(c) portrays the Precision-Recall and Precision graph, where precision designates a test boundary that sanctions us to evaluate the all-out number of estimates a model is equipped for making. Equations 1, 2, and 3 are habituated to calculate the precision, precision, and recall values, respectively. Utilizing equations 1, 2, and 3, we get, Precision score is 96.2303% (up to 4 decimal numbers), a Precision score is 96.6389% (up to 4 decimal numbers), and a Recall score is 93.5907% % (up to 4 decimal number) independently shown in table 1. This can ascertain the estimations are working adequately.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \text{----- (1)}$$

$$\text{Precision} = \frac{TP}{TP + FP} \text{..... (2)}$$

$$\text{Recall} = \frac{TP}{TP + FN} \text{..... (3)}$$

Where TP indicates True Positive, TN means True Negative, FP signifies False Positive, FN means False Negative.

#### 4. CONCLUSION

The issue cognate to spam is elongating significantly worldwide in this manner torturing virtually every country and isn't at all appearance the denotement of moving down since the number of people, similarly like the number of online organizations at unobtrusive rates, are moreover growing. With elongated globalization and elongated capacity to verbalize unreservedly, the demonstration of mishandling them as spam comments on sundry progressed stages where people can voice everything without exception has been on the ascending as well. Congruously. Lately, a broad scope of techniques has been proposed for assessment spam identification. In this paper, a robust SVM strategy has been proposed to ascertain spam segregating measures utilizing an orchestrated ML-predicated slip into transgression model appearance quality results with mean square error 0.0377 and in chi-square, hypothetical testing offering positive hint optically discerning precision correspondingly as evaluations are likewise estimable. The tribulations' outcomes show that this strategy gives copacetic records when contrasted with those dependent on most quantifications.

#### References

- [1] R. Kaur, S. Singh, and H. Kumar, "Rise of spam and compromised accounts in online social networks: A state-of-the-art review of different combating approaches," *J. Netw. Comput. Appl.*, vol. 112, pp. 53–88, 2018, doi: 10.1016/j.jnca.2018.03.015.
- [2] H. Wang, Y. Li, and K. Guo, "Countering web spam of link-based ranking based on link analysis," *Procedia Eng.*, vol. 23, pp. 310–315, 2011, doi: 10.1016/j.proeng.2011.11.2507.
- [3] A. Fronzetti Colladon and P. A. Gloor, "Measuring the impact of spammers on e-mail and Twitter networks," *Int. J. Inf. Manage.*, vol. 48, no. April 2017, pp. 254–262, 2019, doi: 10.1016/j.ijinfomgt.2018.09.009.
- [4] H. Shen, "Neurocomputing Discovering social spammers from multiple views," *Neurocomputing*, vol. 225, no. March 2016, pp. 49–57, 2017, doi: 10.1016/j.neucom.2016.11.013.
- [5] J. Yin, Q. Li, S. Liu, Z. Wu, and G. Xu, "Leveraging multi-level dependency of relational sequences for social spammer detection," *Neurocomputing*, vol. 428, pp. 130–141, 2021, doi: 10.1016/j.neucom.2020.10.070.
- [6] S. Juan Ji et al., "A burst-based unsupervised method for detecting review spammer groups," *Inf. Sci. (Ny)*, vol. 536, pp. 454–469, 2020, doi: 10.1016/j.ins.2020.05.084.
- [7] Q. Fu, B. Feng, D. Guo, and Q. Li, "Combating the evolving spammers in online social networks," *Comput. Secur.*, vol. 72, pp. 60–73, 2018, doi: 10.1016/j.cose.2017.08.014.
- [8] C. A. Shue, M. Gupta, C. H. Kong, J. T. Lubia, and A. S. Yuksel, "Spamology: A study of spam origins," 6th Conf. Email Anti-Spam, CEAS 2009, no. January, 2009.
- [9] S. Hedley, "A brief history of spam," *Inf. Commun. Technol. Law*, vol. 15, no. 3, pp. 223–238, 2006, doi: 10.1080/13600830600960758.
- [10] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: A survey," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3634–3642, 2015, doi: 10.1016/j.eswa.2014.12.029.
- [11] S. P. Williams and V. Hausman, "Categorizing the Business Risks of Social Media," *Procedia Comput. Sci.*, vol. 121, pp. 266–273, 2017, doi: 10.1016/j.procs.2017.11.037.
- [12] N. Saidani, K. Adi, and M. S. Allili, "A semantic-based classification approach for an enhanced spam detection," *Comput. Secur.*, vol. 94, p. 101716, 2020, doi: 10.1016/j.cose.2020.101716.
- [13] T. Oladipupo, "Machine Learning Overview," *New Adv. Mach. Learn.*, 2010, doi: 10.5772/9374.
- [14] S. Gupta, J. Sarkar, M. Kundu, N. R. Bandyopadhyay, and S. Ganguly, "Automatic recognition of SEM microstructure and phases of steel using LBP and random decision forest operator," *Meas. J. Int. Meas. Confed.*, vol. 151, p. 107224, 2020, doi: 10.1016/j.measurement.2019.107224.
- [15] S. Gupta et al., "Modelling the steel microstructure knowledge for in-silico recognition of phases using machine learning," *Mater. Chem. Phys.*, vol. 252, no. March, p. 123286, 2020, doi: 10.1016/j.matchemphys.2020.123286.

- [16] S. Panda, A. K. Ghosh, A. Das, U. Dey, and S. Gupta, "Machine Learning-based Linear regression way to deal with making data science model for checking the sufficiency of night curfew in Maharashtra , India," vol. 1, no. 2, pp. 168–173, 2021.
- [17] M. Abbaszadeh, A. Hezarkhani, and S. Soltani-Mohammadi, "An SVM-based machine learning method for the separation of alteration zones in Sungun porphyry copper deposit," *Chemie der Erde*, vol. 73, no. 4, pp. 545–554, 2013, doi: 10.1016/j.chemer.2013.07.001.
- [18] A. M. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, "Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts," *Knowledge-Based Syst.*, vol. 153, pp. 91–104, 2018, doi: 10.1016/j.knosys.2018.04.025.
- [19] H. Lattar, A. Ben Salem, and H. H. Ben Ghezala, "Does data cleaning improve heart disease prediction?," *Procedia Comput. Sci.*, vol. 176, pp. 1131–1140, 2020, doi: 10.1016/j.procs.2020.09.109.
- [20] Y. Huang, M. Milani, and F. Chiang, "Privacy-aware data cleaning-as-a-service," *Inf. Syst.*, vol. 94, p. 101608, 2020, doi: 10.1016/j.is.2020.101608.
- [21] D. C. Corrales, A. Ledezma, and J. C. Corrales, "A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks," *Appl. Soft Comput. J.*, vol. 90, p. 106180, 2020, doi: 10.1016/j.asoc.2020.106180.
- [22] C. M. Yeomans, R. K. Shail, S. Grebby, V. Nykänen, M. Middleton, and P. A. J. Lusty, "A machine learning approach to tungsten prospectivity modelling using knowledge-driven feature extraction and model confidence," *Geosci. Front.*, vol. 11, no. 6, pp. 2067–2081, 2020, doi: 10.1016/j.gsf.2020.05.016.
- [23] J. Wang and M. Wang, "Review of the emotional feature extraction and classification using EEG signals," *Cogn. Robot.*, vol. 1, no. December 2020, pp. 29–40, 2021, doi: 10.1016/j.cogr.2021.04.001.
- [24] M. Maeder, N. McCann, S. Clifford, and G. Puxty, *Model-Based Data Fitting*, Second Edi., vol. 3, no. January. Elsevier, 2020.
- [25] O. Ozturk Mizrak, C. Mizrak, A. Kashkynbayev, and Y. Kuang, "Can fractional differentiation improve stability results and data fitting ability of a prostate cancer model under intermittent androgen suppression therapy?," *Chaos, Solitons and Fractals*, vol. 131, no. xxxx, p. 109529, 2020, doi: 10.1016/j.chaos.2019.109529.
- [26] M. Guo and M. Ghosh, "Mean squared error of James-Stein estimators for measurement error models," *Stat. Probab. Lett.*, vol. 82, no. 11, pp. 2033–2043, 2012, doi: 10.1016/j.spl.2012.06.019.
- [27] J. F. Quessy, L. P. Rivest, and M. H. Toupin, "Goodness-of-fit tests for the family of multivariate chi-square copulas," *Comput. Stat. Data Anal.*, vol. 140, pp. 21–40, 2019, doi: 10.1016/j.csda.2019.04.008.



**Priyanka Dhara** (Pursuing M.C.A. Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)



**Sohom Bhattacharya** (Pursuing M.C.A. Dr. B.C Roy Engineering College Durgapur, West Bengal – 713206, India)



**Shubham Bhattacharjee** (Pursuing M.C.A. Dr. B.C Roy Engineering College Durgapur, West Bengal – 713206, India)



**Dr Subir Gupta** (Assistant Prof MCA department Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)



BEST  
BCREC Engineering & Science Transaction  
Journal website: [www.bcrecjournal.org](http://www.bcrecjournal.org)



## ML-BASED SURVEY USING LINEAR EQUATION

Ankita Talukdar<sup>1</sup>, Debasis Mandal<sup>2</sup>, Uttam Dey<sup>3</sup>, Biswajit Mondal<sup>4\*</sup>

<sup>1</sup>Department of Faculty and Management Studies, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India, Email: [mimiwinxo@gmail.com](mailto:mimiwinxo@gmail.com)

<sup>2</sup>Dept of Computer Science, IMPS College Of Engineering And Technology, West Bengal – 732103, India, Email: [debimps07@gmail.com](mailto:debimps07@gmail.com)

<sup>3</sup>Academy of Professional Courses, Dr. B. C. Roy Engineering Coll, Durgapur, West Bengal – 713206, India, Email: [uttam.de@bcrec.ac.in](mailto:uttam.de@bcrec.ac.in)

<sup>4\*</sup> Department of CSE, Dr. B. C. Roy Engineering Coll, Durgapur, West Bengal – 713206, India, Email: [biswajit.mondal@bcrec.ac.in](mailto:biswajit.mondal@bcrec.ac.in)

**Abstract-** Staying indoors proved to be the only solution since the fundamental question of survival hovered over the Homo sapiens for more than a year and a half unless the vaccination was manufactured to combat the super-mutating virus. In December 2019, the first case of Coronavirus was first witnessed. The root of the problem of the Pandemic is believed to have taken birth in the city of Wuhan, China. The virus had taken all over the globe and people helplessly witnessed countless deaths and sufferings. A century since the last pandemic, people are going through the toughest time quarantining themselves and distancing themselves from one another. The second wave had shown India the most frightful scenario with positive cases rising like flames and scarcity of oxygen, hospitals and so many other necessities. The health, economy, and similar arenas received a massive blow due to the novel Coronavirus. Despite repeated warnings and alerts, the death and infected ratio in India knew no height during the second wave.

With fewer facilities and apt medicines, two consecutive waves of Covid had created an upheaval among the citizens of India. This paper displays the heavy usage of Data Science (DS), Machine Learning (ML), and Linear Equation to bring into view the details of the virus' effect in India. The study done on the gruesome condition includes three-time phases from the initial stage of the Pandemic till 30<sup>th</sup> June 2021. The first period is from 25/03/2020 to 30/09/2020, the second being from 01/10/2020 to 31/03/2021 and the final period was from 01/04/2021 to 30/06/2021. The prime motif embedded in this paper is to spill the various steps and measurements implemented by the states of India that turned into failure in wrestling the immensely powerful virus and at the same time making one know what to avoid in the upcoming time.

**Keywords-** Covid-19; Data Science; Linear Equation; Machine Learning; saddle point.

\*Corresponding author: Biswajit Mondal, [biswajit.mondal@bcrec.ac.in](mailto:biswajit.mondal@bcrec.ac.in)

## 1. INTRODUCTION

The Spanish Influenza in 1918 preceded Covid-19, nearly a century ago and delineated the severely infected condition of millions summing up to about one-third of the world's population then[1][2][3]. It's been more than a year since the first case was reported that people have been witnessing the infection spreading like wildfire scenarios and locking themselves to be safe. With 'Social distancing' becoming the slogan, it is a struggle both inside and out[4][5]. Both the waves have equally proven fatal globally and cleared a shocking percentile of the population. Several methods have been applied by countries all over the world to fight against the deadly virus, including India. With no appropriate medicine in hand, the basic step that is to be taken is the absolute dissolving of human engagement.

The lesser the gathering, the least radiation of the disease[6]. So, the immediate lockdown of International flights, transport systems, shopping malls, schools, colleges, workplaces, and every possible place of interaction on a large scale was the first and foremost step opted globally. Guidelines have been developed and passed on socially to ensure minimum dispersal of this deadly disease. The pandemic proved not only to be heavy upon the population count but also stroke badly down the economy. The 32 states of India might have shielded themselves during the first wave prettily, but the second made a frightful impact with the picture of the scarcity of hospitals, services, oxygen, and other necessities everywhere[7][8]. The daily death count for sure gave the uninfected ones goosebumps and also the message of following the rules of safety strictly. This paper thus proves to help undrape the methods that became either winners or failures to be prepared for the time ahead.

Data Science, Machine Learning, Artificial Intelligence are the few areas that are commonly spoken about in daily life[9][10][11][12][13][14][15]. Apart from making regular life pretty easy and energy-saving, the impact of these in the field of science is hugely appreciable. Machine Learning or ML can be defined as an Application backed by AI that helps a system to develop its ability to learn on its own from previously acquired experience without having one to tutor or program[16][17][18][19]. It is purely gorgeous that the system identifies the pattern and utilizes and functions accordingly. Several systems can be studied, but the current paper has opted linear regression model to examine the ongoing issue[20][21].



## 2. METHODOLOGY

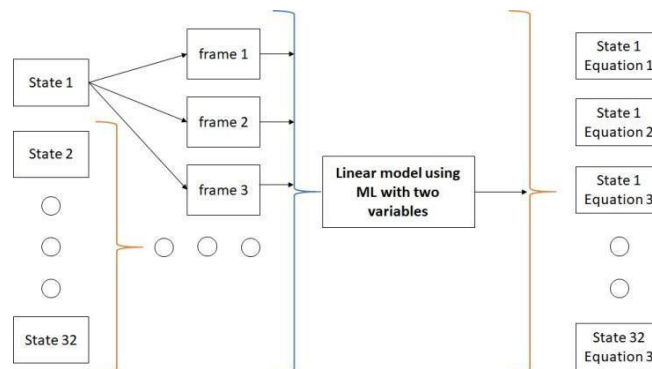


Figure 1: Model construction using a linear model with two variables

It is of no doubt that the effect of the Corona virus is unparalleled. Just like the condition is not the same when two countries are compared, so can be stated regarding the 32 states of India. While one might think that Maharashtra might be the worst sufferer, studies might not give a nod at this. The current paper uses a method to detect the condition of every state and also find out the methods that have been implemented have proved to be positive or not. Quite a few steps are involved in solving the puzzle of finding out the states that successfully battled against the virus by following certain guidelines and those states who turned to be the red zone for having the highest number of cases and deaths. The process followed while carrying out the study involves the inclusion of Linear regression; multiple tables and graphs are present to give a better view in studying the condition of the states.

The collection of data from the government API and also private API is the first step of examining the states' condition thoroughly[22]. The data received is in bulk and has to be thus filtered to extract only the required portion. As stated earlier that the entire study has been done by diving the pandemic into three time frames from 25/03/2020 to 30/06/2020. The first-time frame is from 25/03/2020 to 30/09/2020, the second time frame is from 01/10/2021 to 31/03/2020, and the final is from 01/04/2021 to 30/06/2021. The analysis of the filtered data is the next step which is done by training a model, that is, ML-based data science model, which is fed with data, and Linear regression method is used.

## 3. RESULT ANALYSIS

Studying or testing about the novel virus without any doubt is a very tough subject of the contemporary. Ranging from collecting data, studying about the virus that is mutating over and over again to the grievous condition of the globe, shattering under its demonic grasp, new theories and conclusions of the virus are emerging regularly. India, being an immensely vast country, is relatively difficult to study. As the aim of the Paper is to detect the condition of each state of the country, hence carefully data was found separately. The tables show the condition of each state in three-time frames. The equation developed  $dy/dx$  have clearly shown the performance, how different they have been in the first phase, second and third. The lesser the value, the greater their shift towards green zone, whereas when the values have exceeded, straight away they fall into the danger zone. Each time frame's result can state the success of the measures taken by each state, that is, whether they have been capable enough to secure the people from the deadly virus or not. The reason behind the

differing of the values solely depend upon the safety measures implemented or the scale of carelessness. We definitely know that the second wave's hit had proved horribly fatal, hence the performance of every state as per the results can be taken into consideration and be prepared for the times ahead. Hence the Linear equation usage when summing up the result, 96 quadratic equations has been formed comprising three from each state.








































While the motive is to find out the condition of being affected by the virus using the model, an equation has been established during the procedure:

Using equation (1) we find the values of  $dy/dx$  and constant of each frame

$$L(t) = 1 / (1 + e^{-t}) \text{ Where } t = a + bx \text{----- (1)}$$

Table 1: State wise data

State Name	1 <sup>st</sup> frame		2 <sup>nd</sup> frame		3 <sup>rd</sup> frame	
	dy/dx	Constant	dy/dx	Constant	dy/dx	Constant
Andaman	0.004	-692.101	0.002	4034.623	0.007	-6238.149
Arunachal	0.018	-2261.74	0.01	56852.809	0.022	-29047.81
Assam	0.067	-21057.824	0.04	498513.65	0.071	-216828.42
Bihar	0.071	-23417.54	0.042	39526.894	0.07	-238421.947
Chandigarh	0.016	-1264.27	0.01	28923.068	0.019	-12816.812
Chhattisgarh	0.059	-1842.529	0.037	36984.205	0.064	-18354.686
Daman & Diu	0.004	-527.16	0.002	7268.681	0.006	-5964.279
New Delhi	0.108	-26042.375	0.071	529610.54	0.113	-248105.61
Goa	0.029	-1448.665	0.02	22978.78	0.034	-14355.094
Gujrat	0.054	-1794.824	0.021	4695.035	0.067	-17428.492
Haryana	0.046	-16358.167	0.012	305495.206	0.051	-167245.484
HimachalPradesh	0.019	-6406.59	0.01	8692.601	0.023	-68429.88
Jammu&Kashmir	0.037	-9274.354	0.011	109523.26	0.043	-95713.813
Jharkhand	0.04	-429.046	0.002	6209.315	0.07	-4903.731
Karnata	0.174	-32440.532	0.081	672969.318	0.194	-360421.394
Kerala	0.114	-27692.827	0.094	338415.307	0.132	-270934.892
Ladakh	0.008	-1689.492	0.005	269520.85	0.011	-16729.978
Lakshadweep	0.002	-256.847	0.001	6854.29	0.008	-2569.492
Madhya Pradesh	0.051	-1763.926	0.019	220571.282	0.0067	-17543.831
Maharashtra	0.204	-41091.349	0.088	916386.788	0.271	-413767.438
Manipur	0.017	-1169.407	0.009	29387.049	0.02	-11386.724
Meghalaya	0.012	-894.754	0.008	1025.924	0.019	-838.245
Mizoram	0.007	-418.874	0.005	1327.275	0.011	-12145.786
Nagaland	0.009	-962.47	0.006	1096.254	0.013	-9284.15
Odisha	0.081	-24217.341	0.018	46201.258	0.09	-24861.957
Puducherry	0.026	-14127.628	0.009	269842.047	0.031	-146825.285
Punjab	0.043	-1557.462	0.023	25463.29	0.052	-15927.256
Rajasthan	0.064	-19402.781	0.04	22629.137	0.073	-195896.45
Sikkim	0.003	-562.78	0.001	8624.573	0.008	-5628.673
Tamil Nadu	0.143	-29651.943	0.099	325981.082	0.16	-296842.791
Telengana	0.076	-23604.856	0.029	26248.397	0.082	-23968.217
Tripura	0.023	-1928.194	0.017	21269.568	0.29	-19867.382
Uttarakhand	0.034	-11684.273	0.019	17286.04	0.041	-11698.249
Uttar Pradesh	0.121	-28471.647	0.081	329272.068	0.166	-289682.52
West Bengal	0.188	-24670.16	0.054	136800.295	0.113	-789029.35

State Name	1 <sup>st</sup> frame	2 <sup>nd</sup> frame	3 <sup>rd</sup> frame
Andaman and Nicobar Islands			
Arunachal Pradesh			
Andhra Pradesh			
Assam			
Bihar			
Chandigarh			
Chattisgarh			
Dadra and Nagar Haveli and Daman and Diu			
New Delhi			
Goa			
Gujrat			
Haryana			
Himachal Praesh			



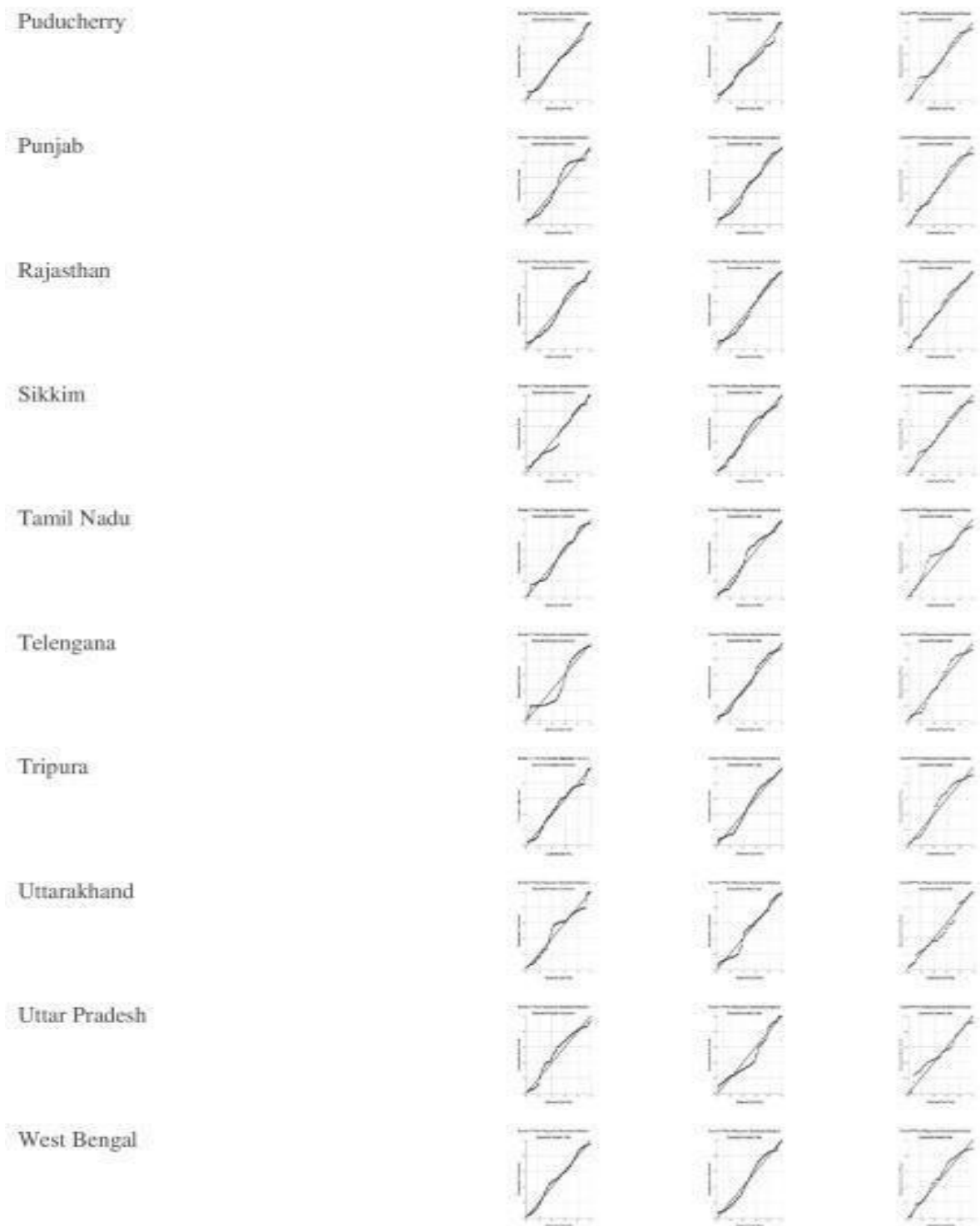


Figure 2: State wise linear graph

#### 4. CONCLUSION

The novel Coronavirus had not only been in the headline for quite some time now but also the most spoken about be it the Google search engine or in a presentation showing its features or effect. India, one of the heavily populated country had seen the harsh days during the second wave for not taking enough steps. The frequently mutating virus could not be combatted until the vaccine was

manufactured. Still, lockdown being the vital solution to control further loss of life, the percentile of infected and recovered might have seen some changes recently after the vaccination process has come into the limelight. The implementation of the methods in states, as per the values mentioned in the table is highly appreciable for the infected ratio is in decent shape. On the other hand, the ones with maximum values have failed for neglecting the guidelines and ended up having a frightful percentile of deaths and the infected. The prime objective of this paper was not only to find out the states' name and their methods used but also to learn about these methods and keep in mind that they could be useful in the times ahead if the wave is seen. However, the success lies in using Machine Learning to detect the condition, but few points have to be kept in mind too. Only a few parameters have been given the stage during the survey, while several others were missed, namely population, bed count, the count of oxygen cylinders, area of the state and more similar facilities. In order to go further with these parameters, such model is not enough to achieve satisfactory results and has to be replaced by higher, complex model. When these parameters are utilized during the survey is done, the complete scenario can be established consisting the loops and holes formed while adopting the safety measures, for it is absolutely necessary to learn about all those steps which have contributed towards the superlative performance.

## 5. Acknowledgment

The authors sincerely acknowledge the expert guidance received during authoring the present work from Prof. (Dr.) Subir Gupta, Department of Computer Science & Engineering, Dr. B. C. Roy Engineering College, Durgapur.

## References

- [1] I. H. Sarker, "CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks," *Internet of Things*, vol. 14, p. 100393, Jun. 2021, doi: 10.1016/j.iot.2021.100393.
- [2] E. Kondilis *et al.*, "The impact of the COVID-19 pandemic on refugees and asylum seekers in Greece: A retrospective analysis of national surveillance data from 2020," *EclinicalMedicine*, vol. 37, p. 100958, 2021, doi: 10.1016/j.eclinm.2021.100958.
- [3] R. K. Rajeesh, A. M, B. E, S. J. P. J, K. A, and P. S, "Detection and monitoring of the asymptomatic COVID-19 patients using IoT devices and sensors," *Int. J. Pervasive Comput. Commun.*, 2020, doi: 10.1108/IJPCC-08-2020-0107.
- [4] W. Ding, J. Nayak, H. Swapnarekha, A. Abraham, B. Naik, and D. Pelusi, "Fusion of intelligent learning for COVID-19: A state-of-the-art review and analysis on real medical data," *Neurocomputing*, vol. 457, pp. 40–66, 2021, doi: 10.1016/j.neucom.2021.06.024.
- [5] S. Bhardwaj, S. Sharma, and R. Bhardwaj, "Machine Learned Hybrid Gaussian Analysis of COVID-19 Pandemic in India," *Results Phys.*, p. 104630, Aug. 2021, doi: 10.1016/j.rinp.2021.104630.
- [6] M. Akcay, D. Etiz, and O. Celik, "Prediction of Survival and Recurrence Patterns by Machine Learning in Gastric Cancer Cases Undergoing Radiation Therapy and Chemotherapy," *Adv. Radiat. Oncol.*, vol. 5, no. 6, pp. 1179–1187, 2020, doi: 10.1016/j.adro.2020.07.007.
- [7] V. Cicek, "Corrosion Prevention and Protection," in *Cathodic Protection*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, pp. 97–125.
- [8] H. Zhu, B. Yan, Y. Xu, J. Guan, and S. Liu, "Removal of nitrogen and COD in horizontal

- subsurface flow constructed wetlands under different influent C/N ratios,” *Ecol. Eng.*, vol. 63, pp. 58–63, Feb. 2014, doi: 10.1016/j.ecoleng.2013.12.018.
- [9] A. K. Ghosh, S. Panda, A. Das, U. Dey, and S. Gupta, “Machine Learning-Based Data Science Model for Examination for Corona Virus Second Stage Spread Rate: A Contextual Investigation Utilizing West Bengal Dataset Till 15th May 2021,” *Eng. Technol. J. Res. Innov.*, vol. III, no. II, pp. 35–39, 2021.
- [10] S. Gupta, J. Sarkar, A. Banerjee, N. R. Bandyopadhyay, and S. Ganguly, “Grain Boundary Detection and Phase Segmentation of SEM Ferrite–Pearlite Microstructure Using SLIC and Skeletonization,” *J. Inst. Eng. Ser. D*, vol. 100, no. 2, pp. 203–210, Oct. 2019, doi: 10.1007/s40033-019-00194-1.
- [11] S. Gupta *et al.*, “Modelling the steel microstructure knowledge for in-silico recognition of phases using machine learning,” *Mater. Chem. Phys.*, vol. 252, no. May, p. 123286, Sep. 2020, doi: 10.1016/j.matchemphys.2020.123286.
- [12] S. Panda, A. K. Ghosh, A. Das, U. Dey, and S. Gupta, “Machine Learning-based Linear regression way to deal with making data science model for checking the sufficiency of night curfew in Maharashtra , India,” *Int. J. Eng. Appl. Phys.*, vol. 1, no. 2, pp. 168–173, 2021.
- [13] B. Mondal, “Artificial Intelligence: State of the Art,” in *Intelligent Systems Reference Library*, vol. 172, 2020, pp. 389–425.
- [14] S. Bhattacharya, S. Bhattacharjee, A. Das, A. Mitra, and I. Bhattacharya, “Machine learning-based Naive Bayes approach for divulgence of Spam Comment in Youtube station,” vol. 1, no. 3, pp. 278–284, 2021.
- [15] S. Gupta, “Chan-Vese segmentation of SEM ferritepearlite microstructure and prediction of grain boundary,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 10, pp. 1495–1498, Aug. 2019, doi: 10.35940/ijitee.A1024.0881019.
- [16] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.
- [17] Z. Zhu, Y. Wang, and G. Jiang, “Unsupervised segmentation of natural images based on statistical modeling,” *Neurocomputing*, vol. 252, pp. 95–101, 2017, doi: 10.1016/j.neucom.2016.03.117.
- [18] Z. H. Zhou, N. V. Chawla, Y. Jin, and G. J. Williams, “Big data opportunities and challenges: Discussions from data analytics perspectives [Discussion Forum],” *IEEE Computational Intelligence Magazine*, vol. 9, no. 4. Institute of Electrical and Electronics Engineers Inc., pp. 62–74, Nov. 01, 2014, doi: 10.1109/MCI.2014.2350953.
- [19] T. Oladipupo, “Machine Learning Overview,” *New Adv. Mach. Learn.*, 2010, doi: 10.5772/9374.
- [20] B. T. Pham and I. Prakash, “Evaluation and comparison of LogitBoost Ensemble, Fisher’s Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping,” *Geocarto Int.*, vol. 34, no. 3, pp. 316–333, 2019, doi: 10.1080/10106049.2017.1404141.
- [21] P. K. M. R. Palanivel 1, “Prediction and optimization of process parameter of friction stir welded AA5083? H111 aluminum alloy using response surface methodology R.,” *J. Cent. South Univ. 19 1–8*, vol. 19, pp. 1–8, 2012, doi: 10.1007/s11771.

- [22] R. Alfred, "The rise of machine learning for big data analytics," no. February, pp. 1–1, 2017, doi: 10.1109/icsitech.2016.7852593.

### BIOGRAPHIES OF AUTHORS

	<p>Ankita Talukdar Department of FMS Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India E-Mail: mimiwinxo@gmail.com</p>
	<p>Debasis Mandal Dept of Computer Science IMPS College Of Engineering And Technology Malda, West Bengal –732103, India Email: debimps07@gmail.com</p>
	<p>Ayan Kumar Ghosh Academy of Professional Courses Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India Email: ayanghosh63@gmail.com</p>
	<p>Biswajit Mondal* Department of CSE Durgapur, West Bengal –713206, India Email: biswajit.mondal@bcrec.ac.in *Corresponding Author</p>





# RAINFALL RUNOFF MODELLING OF THE KUSHKARANIKA RIVER CATCHMENT, BIRBHUM, WEST BENGAL, INDIA USING ARTIFICIAL NEURAL NETWORK

Arnab Mitra<sup>1</sup>, Debaduyti Nath<sup>2</sup>

<sup>1</sup>Department of Civil Engineering, Abacus Institute of Engineering & Management, Natungram Magra, West Bengal – 712148, India,  
Email: [arnabmitraiest@gmail.com](mailto:arnabmitraiest@gmail.com)

<sup>2</sup>Department of Mining Engineering, Indian Institute of Engineering Science and Technology, Shibpur, West Bengal –711103, India,  
Email: [debaduyti.nath@gmail.com](mailto:debaduyti.nath@gmail.com)

**Abstract** — Rainfall and runoff are major elements of the hydrological cycle, the creation of hydrological structures and the drainage system architecture. In order to determine and foresee its effects, the same is estimated. Direct rainfall-runoff estimation is invariably effective, however, in the time frame required, it is not feasible for most areas. The use of remote sensing and GIS technology can help to solve the problem with traditional runoff estimation methods. In this research, the modified Soil Conservation System (SCS) CN approach is applied to estimate runoff, which takes into account factors such as slope, land use/land cover, and watershed area.

Artificial neural networks (ANNs) are increasingly being used in the study of hydrology and water resources issues. In this research, an ANN was developed and used to model the rainfall-runoff relationship, in Kushkaranika river catchment located in Birbhum, West Bengal, India. The multilayer perceptron (MLP) neural network was chosen for use in the current study. The results of the study show that the artificial neural network method is better for predicting river runoff than the standard regression model. The model is capable of predicting runoff in the future.

Keyword:- *R-R Modelling, Remote Sensing, GIS, SCS-CN, ANN.*

## 1. INTRODUCTION

Flooding happens when a large amount of water is pushed by rivers, creeks, and a variety of other natural features into regions where it cannot be efficiently drained. During periods of excessive rainfall, drainage systems in residential areas are frequently insufficient, or unregulated civil development substantially impairs the performance of a drainage system that is otherwise adequate [1]. Floods harm a big number of people in every country, but in India, because to its extraordinarily high population density and frequently unenforced development norms, a large amount of damage and many deaths that could have been avoided are permitted to occur [2]. Flooding occurs in India as a result of severe rain, which causes rivers, lakes, and dams to overflow, causing significant damage to people's lives and property. In the past, India has experienced several of the world's worst floods, wreaking havoc on people's livelihoods, property, and critical infrastructure [10].

Rainfall, in particular, has a significant impact on agriculture. Because all plants require at least some water to survive, agriculture relies heavily on rain (which is the most efficient method of watering). While a consistent rain pattern is necessary for healthy plants, too much or too little rainfall can be damaging to crops, even deadly [3]. Drought can destroy crops and cause erosion, while excessive rain can promote the growth of hazardous fungi. Plants require different quantities of rain to survive. Certain

cactus, for example, require only a tiny amount of water, whereas tropical plants can require hundreds of inches of rain every year to survive.

Modelling the rainfall–runoff processes of hydrology is necessary for a variety of reasons. The fundamental reason is that hydrological measurement techniques have limits. It is impossible to quantify all aspects of hydrological systems. Only a restricted set of measurement techniques and a limited set of spatial and temporal measurements are required for this [2][3][4]. To assess the potential impact of future hydrological change, we need a mechanism to extrapolate from available data in both space and time, especially to ungauged catchments (where measurements are not available) and into the future (where measurements are not possible). Quantitative extrapolation or prediction is possible with different sorts of models, which should be useful in decision-making. A lot of rainfall–runoff modelling is there that is done only for academic reasons to formalise knowledge about hydrological systems [3][4]. Demonstrating such understanding is a critical milestone in the evolution of a scientific area. We learn the most when a model or theory is demonstrated to be in conflict with credible evidence, necessitating some revision of the model's foundational understanding [5]. However, the ultimate goal of employing models to predict hydrological problems must be to improve decision-making in areas such as water resources planning, flood protection, contaminant mitigation, abstraction licencing, and other fields [6]. With increasing demands on water resources around the world, better decision-making in the face of changing weather patterns from year to year necessitates more accurate models [7][8][9]. That is the focus of this study. Rainfall–runoff modelling is done in a purely analytical framework based on measurements of the catchment area's inputs and outputs. The catchment is considered as if it were a black box, with no regard for the internal systems that influence rainfall to runoff conversion [10][21][22]. This model was created in such a way that it was demonstrated that based on an understanding of the nature of catchment response, it may also be feasible to make some physical interpretations of the resulting models. Any rainfall–runoff modelling study must begin with this understanding.

## 2. PROBLEM DEFINITION

The Mayurakshi River consist of four major tributaries, which are Brahmani, Dwarka, Bakreshwar and Kushkaranika and other tributaries are Dhobhai, Bhurburi, Pusaro, Tepra, Sid-heswari, Dauna. The river's upper watershed has a hilly slope topography. By travelling down to the mouth it flows over alluvial plains of Ganga. Kushkaranika River being rain fed is prone to seasonal floods, seven or eight months in the year the river is a desert sands stretching from shore to shore for about 1.5 miles (2.4 km). When the monsoons arrive, however, the river becomes terrifying, demonic, raging over a 4 to 5 mile (6 to 8 km) in width of the channel. It's a river of deep grey water that's engulfing everything. in its path .It is important to note that floods in these areas occurred so frequently that only five years between 1960and 2000 were recorded as flood-free years. But even during this period of time, less than500 square km of area was still deluged. So, by preparing the rainfall-runoff model is needed for this area to see how much surface runoff is coming for any rainfall on a daily basis

## 3. OBJECTIVES OF THE WORK

- To calculate the runoff of the Kushkaranika River Catchment area.
- To generate a model for estimating the runoff generated from any given rainfall data of the similar type of region, using Artificial Neural Network.

- Forecasting and predicting flood peaks, runoff volumes, extreme flood producing rains.

Data collection is one of the most important aspects for conducting any research work. Data collection starts with the planning that what kind of data is required for the research work. The accuracy of the work greatly depends upon the data collection.

#### 4. THEMATIC MAP PREPARATION

##### **STUDY AREA MAP PREPARATION:**

After finalizing the study area for the project work (by the means of field visit, satellite image and Toposheet interpretation) there has been a need of preparing the study area map. At first the toposheet of this study area is georeferenced. The upstream portion of the Kushkaranika River is digitized with the help of ArcGIS 10.5.1 by comparing the river with toposheet and satellite imagery. After digitizing the river the boundary of this river basin became prominent. The mid-point between the tributaries of Kushkaranika River and its adjacent rivers has been selected. These selected points are connected and digitized to get the sub-watershed basin area as shown in the Map.

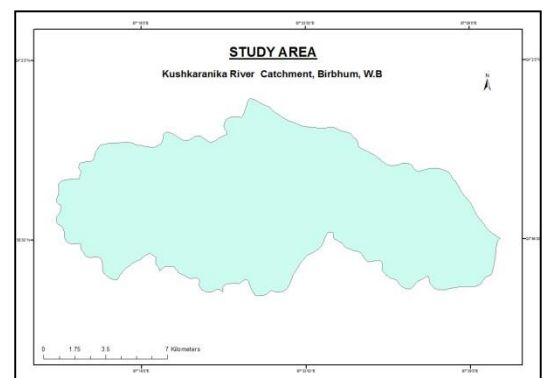


Figure 1:- Study Area

##### **PREPARATION OF SOIL MAP:**

The soil types at a specific location are important because they impact how much water can penetrate the soil and, as a result, how much water flows [11][23]. The soil map was classed based on infiltration capacity, with soil types such as highly infiltrated, moderately infiltrated, and less penetrated being found in the basin. Soil structure and infiltration capacity will have a major impact on the soil's ability to behave absorb water like a sponge. Soil capacity vary depending on the kind of soil. The risk of flooding increases when soil infiltration capacity decreases, resulting in more runoff from the surface. When water is provided at a pace that exceeds the soil's ability to absorb it, it runs off the sloping terrain as runoff, causing flooding [12][21].

For preparation of soil type map, soil maps of different states (Bihar and West Bengal) are collected and georeferenced and extracted. Then each soil features are digitized using a polygon shape file. Each attribute is inserted in the attribute table of digitized vector corresponding to each digitized polygon. Figure shows the extracted soil map.

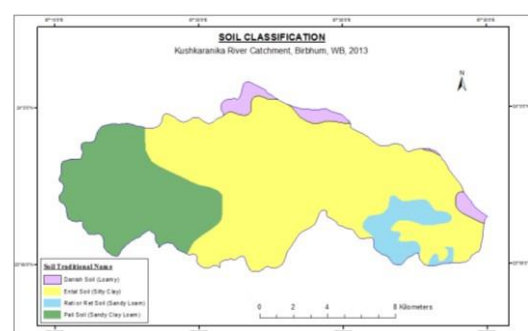


Figure 2:- Soil Map

##### **PREPARATION OF LAND USE LAND COVER (LU-LC):**

One of the most important issues is the area's land use and management, as this is one component that represents not only the current use of the land, its pattern and kind of use, but also the significance of that use in relation to the living population and its interaction with existing development. Soil vegetation

cover, whether permanent grassland or crop cover, has a significant impact on the soil's ability to operate as a water storage reservoir. Rainwater runoff is far more common than on barren fields it is on fields with excellent crop cover. The amount of runoff is reduced when there is a thick vegetation cover [8][10]. Concrete and other impermeable surfaces absorb hardly any water [19]. The research area's land use and land cover classes were created using LISS IV resource satellite image data from NRSC. Using the ERDAS Imagine 2016 programme, a supervised classification method was adopted. The signatures of different pixels present in the satellite image are identified and selected. The necessary signatures are Water, Sand, River, Open Forest, Barren Land, Fallow Land, Deep Forest, Agriculture, Settlement are classified. After that, supervised classification is carried out. Supervised classification is the most used technique for quantitative analysis of remote sensing image data. At its core is the concept of segmenting the spectral domain into parts that can be linked to ground cover classes relevant to a specific application. In classification stage, the Maximum Likelihood Classifier was used to classifying the land use map. The distribution of brightness values are assumed to be normally distributed.

From the prepared first order supervised land use / land cover map, the land use and land cover distribution in the catchment are following:

Table 1:- Land Use/Land Cover Distribution

NO	LU/LC Type	Area (%)	Area (Km <sup>2</sup> )
1	Agriculture	53.05	91.29
2	Water Body	2.52	4.34
3	Forest	21.84	37.58
4	Barren Land	17.51	30.13
5	Settlement	4.42	7.61
6	Sand	0.66	1.14
	<b>Total</b>	<b>100</b>	<b>172.092</b>

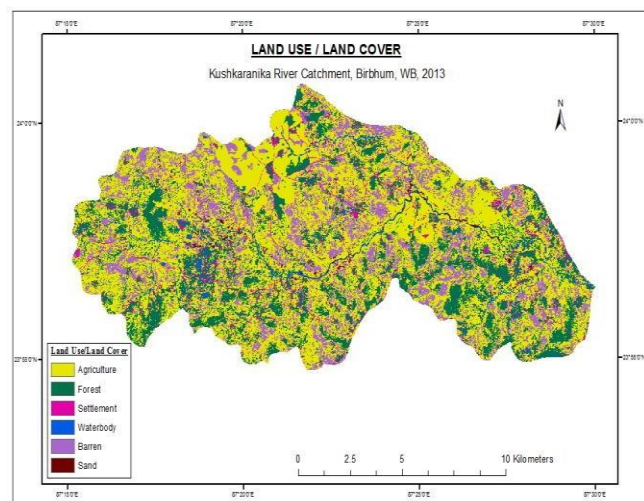


Figure 3:- Land Use/Land Cover Map

**PREPARATION OF RAINFALL DISTRIBUTION MAP:**

Rainfall is the amount of precipitation (rain) in a specified place and time. At a particular place over a definite time period, the rainfall distribution can be represented graphically in the GIS environment. The quantity of precipitation is represented spatially in a distributed manner all over the catchment. The distribution map has been generated working on the principle of Kriging interpolation. The IDW (Inverse Distance Weighted) and Spline interpolation tools are referred to be deterministic interpolation methods since they are directly based on the surrounding observed values or on predefined mathematical rules that determine the smoothness of the final surface [4]. Geostatistical approaches, such as Kriging, belong

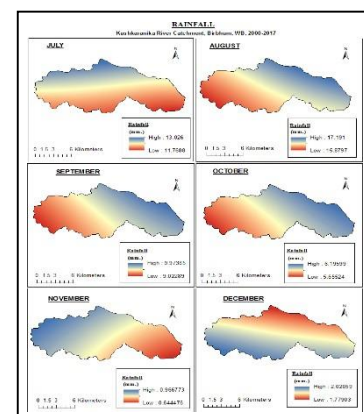
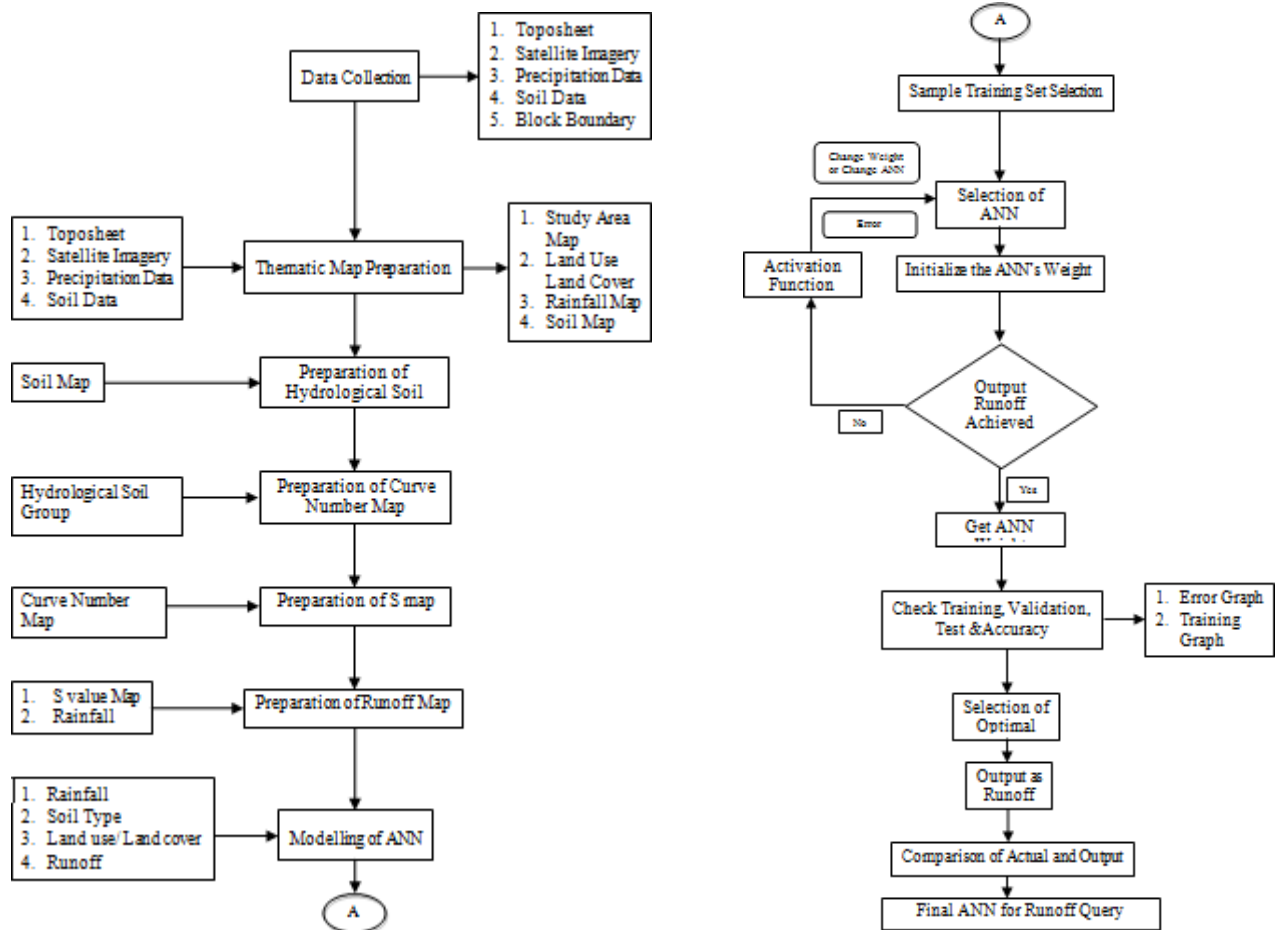


Figure 4:- Rain Fall Distribution Map

to a second type of interpolation approach based on statistical models with autocorrelation—that is, statistical correlations between observed locations [12][19]. As a result, geostatistical methods can produce not only a prediction surface but also a measure of the predictability or accuracy of the predictions [24]. Figure shows the average daily rainfall distribution (2000 - 2017) throughout the catchment for every month.

### 5. METHODOLOGY OF RAINFALL-RUNOFF MODELLING USING



### 6. RESULT AND DISCUSSION

#### **HYDROLOGIC SOIL GROUP (HSG) MAP:**

The distribution soil classes present in the catchment are shown in below table

Sl. No	Land Use/ Land Cover Type	CN Value as per Hydrological Soil Group	
		A	B
1	Agriculture	76	86
2	Water Body	100	100
3	Forest	28	44
4	Barren Land	71	80
5	Settlement	75	75
6	Sand	61	39

From the digitized soil map and above table , it has been found that only Danish Soil (Loamy) lies in the Hydrologic Soil Group “B” and all the other three classes i.e.,Ental Soil (Silty Clay), Rati or Ret Soil (Sandy Loam) and Pali Soil (Sandy Clay Loam) lies in Hydrologic Soil Group “A”.

Table 2:-Hydrological Soil Group Distribution

No.	Soil Type	Area(Km <sup>2</sup> )	% Area
1	Danish Soil (Loamy)	6.302	3.66
2	Ental Soil (Silty Clay)	110.179	64.02
3	Rati or Ret Soil (Sandy Loam)	10.223	5.94
4	Pali Soil (Sandy Clay Loam)	45.388	26.38
	<b>Total</b>	<b>172.092</b>	<b>100</b>

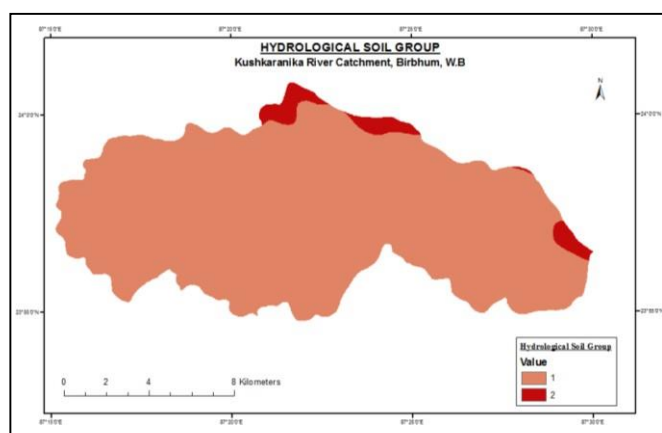


Figure 5:-Hydrological Soil Group Distribution

**CURVE NUMBER MAP:**

By combining hydrologic soil group (HSG) map and land use / land cover map, I have made Curve Number map. The range for Curve Number is 28-100. The opted curve number vales taken from table number- is shown below in the table

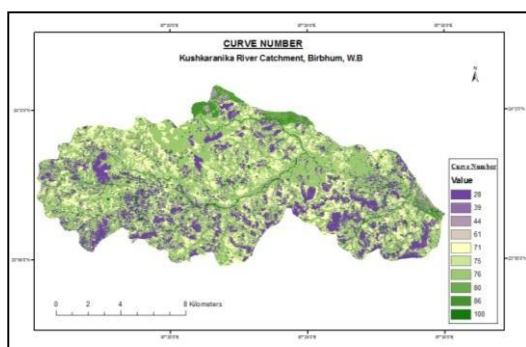


Figure 6:- Curve Number Map

Table 3::- Curve Number Value according to LU/LC Type & Soil Group

No	Soil Type	Area (%)	Area (km <sup>2</sup> )
1	Hydrological Soil Group-A	96.34	165.79
2	Hydrological Soil Group-B	3.66	26.29
	<b>Total</b>	<b>100</b>	<b>172.092</b>

**MAXIMUM POTENTIAL RETENTION MAP:**

From Curve Number Map, I got the Maximum Potential Retention Map by the equation:

$$S = \frac{25400}{CN} - 254,$$

Where, S= Maximum Potential Retention (in mm), CN= Curve Number, a dimensionless number between zero and hundred.

I have used this formula for calculating maximum potential retention map because the unit of rainfall is in millimetre. The range for maximum potential retention is 0-653.

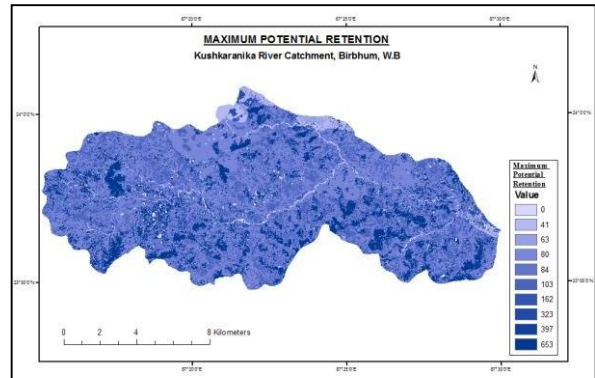


Figure 7:- Maximum Potential Retention Map

**RUNOFF:**

For the determination of runoff by SCS-CN Method, Curve Number Map and Maximum Retention Potential Map have been combined for AMC condition “II”. For determining the runoff, the following formula is been used:

$$V_R = \frac{(P - 0.3 * S)^2}{(P + 0.7 * S)}$$

Where, V<sub>R</sub> = Runoff (in mm), P = Rainfall (in mm), S = Maximum Potential Retention (in mm).

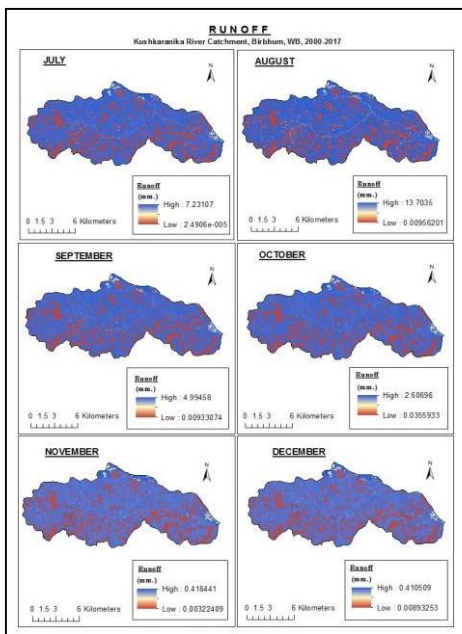


Figure 8:- Runoff Value for month of July to December

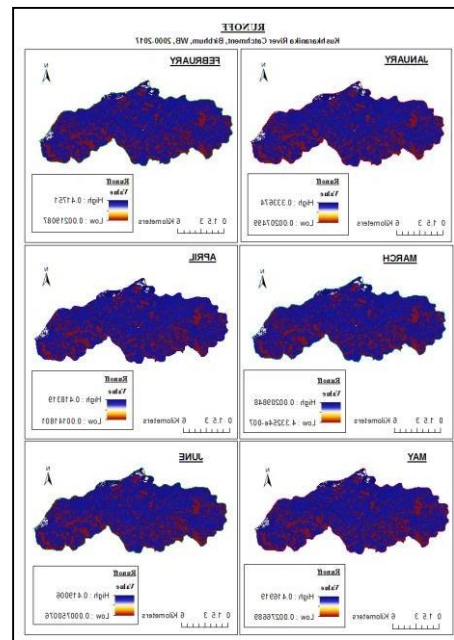


Figure 9:- Runoff Value for month of January to June

After preparing runoff maps by using SCS-CN method following are the results:

**HIGHEST DAILY RAINFALL IN THE 2000-2017 TIME PERIOD:**

The highest daily rainfall of 111.2mm was recorded on 6<sup>th</sup> August.

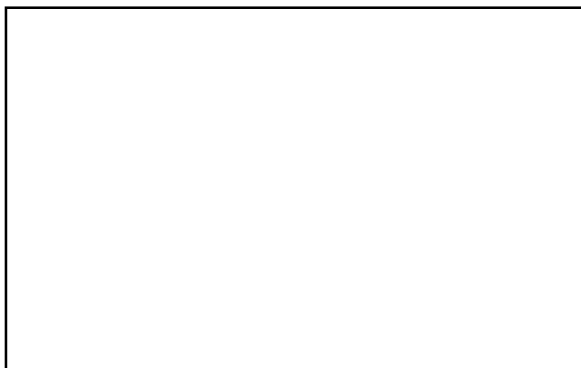


Figure 10:- Highest Daily Rainfall in the time period

**HIGHEST MONTHLY RAINFALL IN THE 2000-2017 TIME PERIOD:**

The highest monthly rainfall recorded was 367.1mm in the month of September 2008.

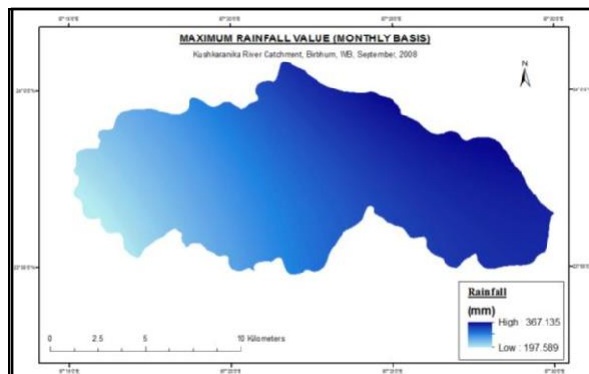


Figure 11:- Highest Monthly Rainfall in the time period

**Highest Yearly Rainfall in the 2000-2017 time period:**

The highest yearly rainfall recorded was 1140mm in the year 2010.



Figure 12:- Highest Yearly Rainfall in the time period

**Minimum observed Runoff in the 2000-2017 time period:**

Minimum runoff of 135mm was observed in the year 2009.



Figure 13:- Minimum Observed Runoff in the time period

**Maximum observed Runoff in the 2000-2017 time period:**

Maximum runoff of 707mm was observed in the year 2007.

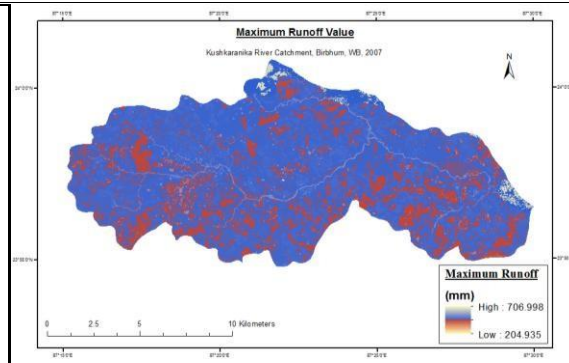


Figure 14:- Maximum Observed Runoff in the time period



### **DESIGN OF ARTIFICIAL NEURAL NETWORK WITH DIFFERENT ARCHITECTURES:**

One input layer, one or two hidden layers, and one output layer were implemented in an MLP ANN architecture. The number of hidden layers was maintained to one or two for the majority of classification problems [14][15][16].

The input layer had three input layers corresponding to the three raster maps and there was only one class in the output as the output layer only contained one neuron. The rainfall at each cell in the study area was being mapped. The value of rainfall varies from 0 to 25.789 representing 0 as the no rainfall condition and 25.789 as the highest rainfall. The intermediate values indicating different rainfall for the Kushkaranika River catchment area.

Trial and error were used to determine the number of neurons in the hidden layer.

### **TRAINING AND EVALUATION OF ANN:**

The weights between the input and hidden layers, as well as between the hidden and output layers, were calculated using the back propagation algorithm [17][18]. MATLAB R2015a software was used to create a three-layered feed-forward network.

To minimize the difference between predicted and calculated output values, the back-propagation technique was applied. The algorithm worked backwards from the error and modified the weights iteratively. The number of iterations was set to 5000, and root mean square error (RMSE) was set to 0.14 as stopping criterion. All the values has been taken as standard value suggested by the software. Validation number was set to 30 as stopping criteria by trial and error validation number has been set.

Variable learning rate, conjugate gradient, and resilient back propagation, as well as Quasi-newton and Levenberg-Marquardt, are all examples of back propagation algorithms. Running the ANN was done using the Levenberg-Marquardt algorithm, which had the fastest convergence [19][20]. The testing dataset was also run through each of the ANN architectures to see how well they could generalise. Gradient descent with momentum weight and bias learning function was employed for the adaptation learning function. The mean squared normalisation error (MSE) was utilised to calculate the performance function. The transfer function was the sigmoid transfer function (TANSIG).

### **ANN ANALYSIS**

The following maps are the input files for creating artificial neural network in MATLAB.

1. Rain fall maps (on a daily basis)
2. Land use / land cover map
3. Soil map
4. Runoff maps (on a daily basis, calculated by using SCS-CN method)

The first 250 data sets have been used for training, the next 58 data sets have been used for testing and the remaining 58 data sets have been used for validation. Then the prepared data sets are processed and comparison between different architectures has been made.

Table 4: Different architectures of ANN

ID	Architecture	Fitness	Train Error	Validation Error	Test Error	Correlation	R-squared
1	3-1-1	3.3080	0.292116	0.279279	0.302294	0.993947	0.986865
2	3-8-1	10.748	0.101263	0.109599	0.093035	0.998789	0.997542
3	3-5-1	12.544	0.097688	0.112746	0.079715	0.998748	0.997453
4	3-3-1	9.2492	0.124741	0.150016	0.108117	0.998314	0.996453
5	3-6-1	11.650	0.098642	0.112604	0.085834	0.998811	0.997576
6	3-4-1	11.233	0.107736	0.107736	0.089019	0.998644	0.997259

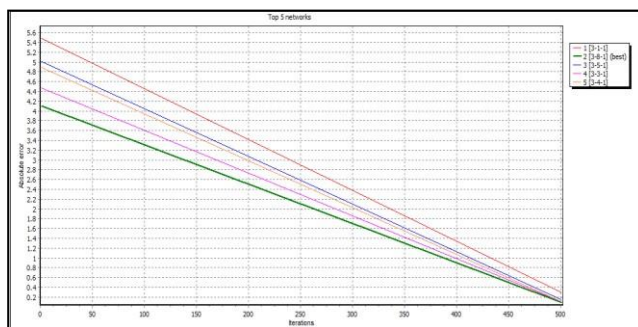


Figure 15: Comparison between top five Networks

All the architectures have been compared for the following categories:

1. Train Error.
2. Validation Error.
3. Test error.
4. Correlation value.
5. R squared value.

By comparing all the results, the best fitting architecture has been modelled as below.

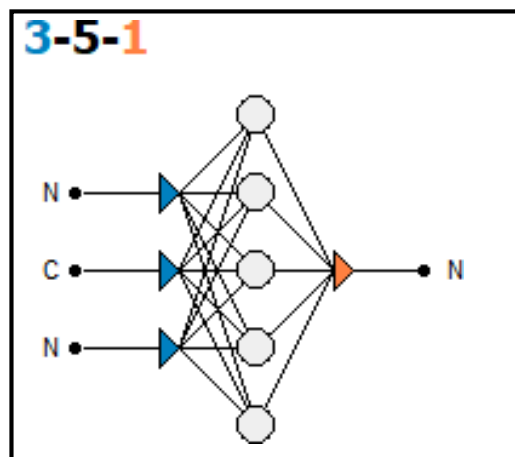


Figure 16: Best Architecture to run the Model

Using the 3-5-1 architecture (where 3 is the number of input, 5 is the number of hidden layer, 1 is the number of output needed from the model) the data sets are trained by using the following criteria:

Training Algorithm –CasCor algorithm (BackPropogation)

Learning Rate – 0.1 (0-100)

Number of Iterations – 5000

After the data sets are trained, the results are produced in the graphical form.

**Generation of Simulated Map:**

After completion of training, simulation has been done by using MATLAB software. For simulation, a well-trained network with the lowest RMSE and an acceptable generalisation difference of error with test data was chosen. The rainfall map on a daily basis, the soil maps which is classified by hydrological soil group and the land use land cover of the area has been is given as input and runoff map is targeted as output.

**Dataset Error Graph:**

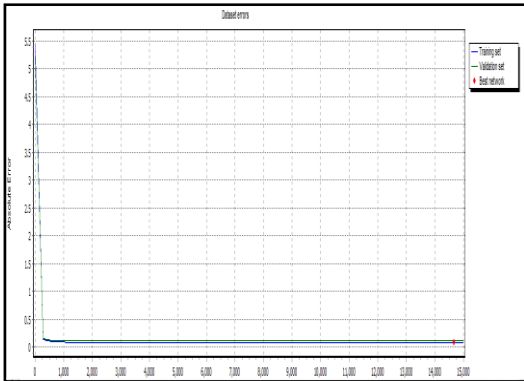


Figure 17:- Dataset error graph

**Error Distribution Graph:**

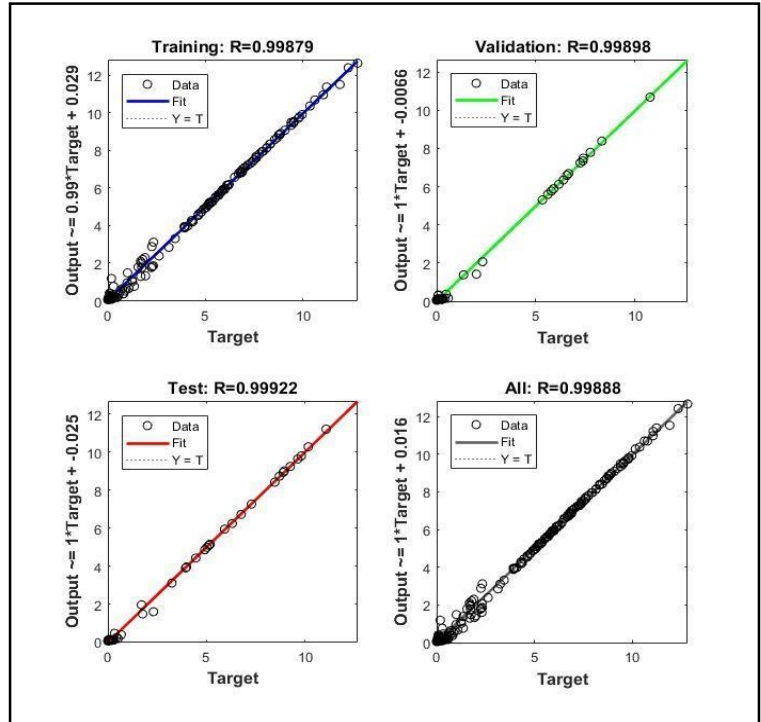


Figure 18:- Error Distribution Diagram

**Error Improvement Graph:**

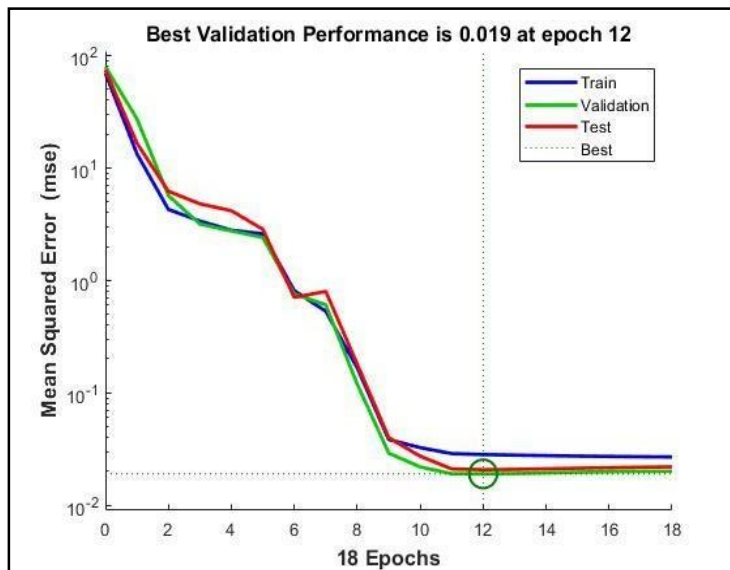


Figure 19:- Error Improvement Diagram

**Result Parameters:**

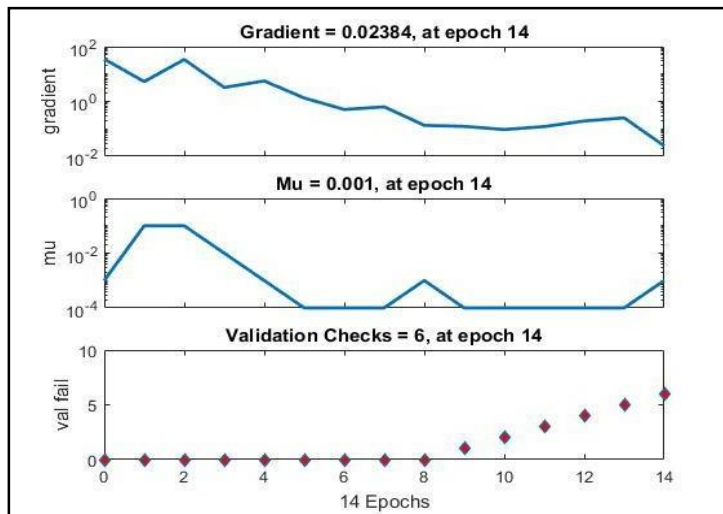


Figure 20:- Result Parameters

Table 5:- Training and Validation of Model Network

	<b>Training</b>	<b>Validation</b>
<b>Absolute Error</b>	0.106176	0.132721
<b>Network Error</b>	0.00017	0
<b>Error Improvement</b>	1.05E-09	
<b>Architecture</b>	3-5-1	
<b>Training Algorithm</b>	Back Propagation	
<b>Training Stop reason</b>	All iterations done	

The model has also been tested internally to see the correlation and for the error parameters too.

Table 6: Mean, standard deviation & minimum value for target

	Target	Output	AE	ARE
Mean	2.303236	2.311323	0.088054	840000275.944313
Standard Deviation	3.312315	3.301218	0.125733	2773878068.05124
Min	0	0.028586	0.000005	6.05E-07
	12.818873	12.647397	1.056042	10000000000

After the testing and validation has been done, the actual runoff values and the runoff values predicted by the model is compared and the final actual versus predicted output graph is given by the model. The model is then ready to be queried.

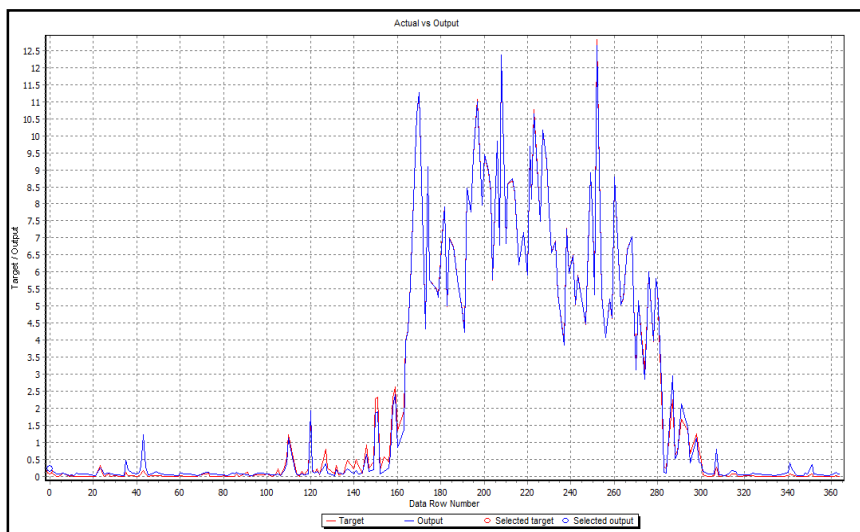


Figure 21:- Actual vs Output Graph

Providing any rainfall value, soil type of that particular point and the land use or land cover type of that area, the ANN model will generate the runoff value.

Some random values are used for testing and validating the model's accuracy. The result for any random value whose catchment area characteristics is same as this will show runoff value per pixel. The rainfall map and runoff map is shown for the maximum rainfall in this year 2018 on 12<sup>th</sup> May.

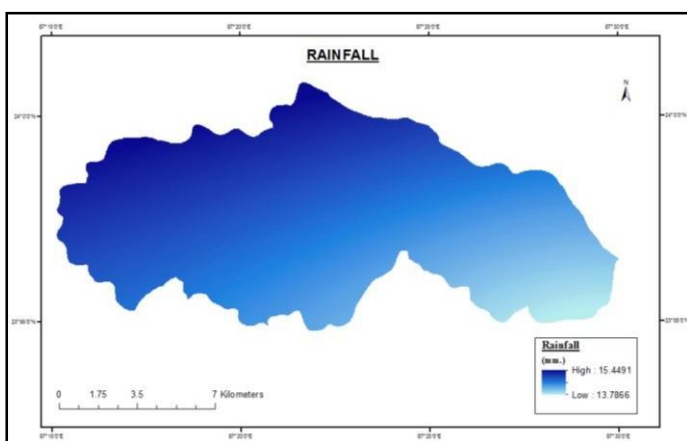


Figure 23:- Rainfall Map of 12th MAY, 2018

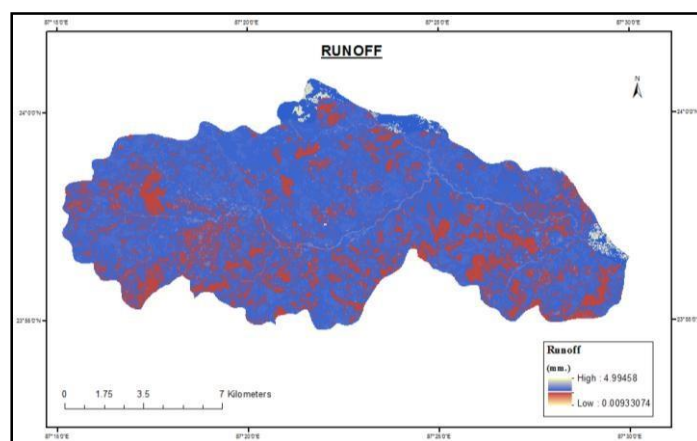


Figure 22:- Runoff Map of 12th May, 2018

## 7. CONCLUSION

The main objective of the project was to calculate the runoff of the Kushkaranika River catchment area, Birbhum, West Bengal, India on a daily basis. So, we can have the idea how much water is coming to the Tilpara barrage as surface runoff. From that we can have the clear idea about scarcity or oversupply of water for irrigation purpose, same thing for hydropower generation also.

For this first SCS-CN method is used in ArcGIS software to calculate the runoff on a daily basis. For calculation of daily runoff in SCS-CN method classified soil map was prepared as per Hydrological Soil Group and Land use/Land cover map was prepared for the study area by collecting the satellite imagery from NRSC with the help of ERDAS Imagine software.

- The runoff for the study area is calculated using SCS-method for a period of 18 years i.e., 2000-2017. The calculated yearly runoff in mm for the years from 2000 to 2017 is 430, 401, 214, 582, 279, 499, 341, 707, 271,135, 372, 476, 289, 384, 296, 459, 510 and 670 mm respectively.
- The runoff for the study area is calculated using SCS-method for a period of 18 years i.e., 2000-2017. The calculated yearly runoff in mm for the years from 2000 to 2017 is 430, 401, 214, 582, 279, 499, 341, 707, 271,135, 372, 476, 289, 384, 296, 459, 510 and 670 mm respectively.
- The monthly runoff and yearly runoff is calculated for the period of 17 years using SCS-CN method. Minimum runoff was observed in the year 2009 and maximum runoff was observed in the year 2007 by using SCS-CN method.
- The daily, monthly, and annual runoff correlation coefficients are 0.73, 0.97, and 0.99, respectively. The yearly runoff graph is better suited than the daily and monthly runoff graphs. Then this runoff data maps were simulated in Matlab for creating a rainfall runoff model by ANN.
- The ability of artificial neural network (ANN) models to model hydrological processes is impressive. In comparison to other traditional models, they are valuable and powerful instruments for dealing with complex situations. The findings reveal that artificial neural networks are capable of modelling rainfall-runoff relationships in places where rainfall and runoff are both fairly regular, validating the overall improvement attained by utilizing neural networks in many other hydrological fields. The artificial neural network (ANN) technique could be a very valuable and accurate tool for solving problems in water resource studies and management.

### **References:-**

1. Birikundavyi S, Labib R, Trung HT, Rousselle J. Performance of neural networks in daily streamflow forecasting. J Hydrol Eng. 2002 Sep; 7(5):392–8.
2. Senthil A.R.K., Sudheer K. P., Jain S. K. and Agarwal P. K. (2004); —Rainfall-runoff modelling using artificial neural networks:comparison of network types. J. Hydrol. Process. 19, 1277–1291.
3. Sobri Harun, and Nor Irwan Ahmat Nor, and Amir Hashim Mohd. Kassim, (2001) Rainfall-Runoff Modelling Using Artificial Neural Network. Malaysian Journal of Civil Engineering, 13 (1). pp. 37-50.

4. Minns AW, Hall MJ. 1996. Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal* 41(3): 399–418.
5. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. II: Hydrologic applications. *J Hydrol Eng ASCE*. 2000 Apr; 5(2):124–37.
6. Anass Boukhris, Stephane Giuliani, Gilles Mourot(2001) ;—Rainfall runoff multi-modelling for sensor fault diagnosis.
7. Campolo M, Andreussi P, Soldati A. River flood forecasting with a neural network model. *Water Resour Res*. 1999; 35(4):1191–7.
8. Patil J.P., Sarangi A., Singh A.K., Ahmad T.(2008); —Evaluation of modified CN methods for watershed runoff estimation using a GIS-based interface.
9. Tayfur G, Singh VP. ANN and fuzzy logic models for simulating event-based rainfall–runoff. *J Hydraul Eng*. 2006 Dec; 132(12):1321–30.
10. Chattopadhyay GS, Choudhury S (2006) Application of GIS and remote sensing for watershed development project – a case study, Map India 2006.
11. Tokar AS, Markus M. Precipitation-runoff modeling using artificial neural networks and conceptual models. *J Hydrol Eng*. 2000 Apr; 5(2):156–61.
12. Subramanya K. (2008); —Engineering Hydrology, Publisher Tata McGraw Hill, 3rd edition, pp. 155-159.
13. Geetha K, Mishra SK, Rastogi AK, Eldho TI, Pandey RP (2005) Identification of dominant runoff generation process using the modified SCS-CN Concept, recent advances in water resources development and management. Nov. 23–25:477–491
14. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. I: Preliminary concepts. *J Hydrol Eng ASCE*. 2000 Apr; 5(2):115–23.
15. Wu J S, Han J, Annambhotla S and Bryant S 2005 Artificial neural networks for forecasting watershed runoff and stream flows; *J. Hydrol. Eng.* 10 216–222.
16. Mishra S., Singh V. (1999)—Behaviour of SCS method in Ia C- Ia- $\lambda$  spectrum pp.112
17. Mishra SK, Singh VP (2002) SCS-CN method: part-I: derivation of SCS-CN based models. *Acta Geophys Pol* 50(3):457–477
18. Kalteh AM. Rainfall-runoff modeling using Artificial Neural Networks (ANNs): Modeling and understanding. *Caspian J Env Sci*. 2008; 6(1):53–8.
19. Rajurkar M.P., Kothiyari U.C., Chaube U.C. (2002); —Artificial neural networks for daily rainfall—runoff modelling. *J. Hydrological Sciences*, 865-877.
20. Chen SM, Wang YM, Tsou I. Using artificial neural network approach for modeling rainfall-runoff due to typhoon. *J Earth Syst Sci*. 2013 Apr; 122(2):399–405.
21. Smith J, Eli RN. Neural network models of rainfall-runoff process. *J Water Resour Plan Manage ASCE*. 1995; 121(6):499–507.
22. Agarwal A, Singh RD, Mishra SK, Bhunya PK. ANN-based sediment yield models for Vamsadhara river basin (India). *Water SA*. 2005; 31(1):95–100.
23. Jain M.K., Mishra S.K. and Singh V.P.(2005) ;—Evaluation of AMC-Dependent SCS-CN-Based Models Using Watershed Characteristics.
24. Atila Dorum , Alpaslan Yarar , M. Faik Sevimli , Mustafa Onucyildiz(2010); —Modelling the rainfall–runoff data of susurluk basin.



**Arnab Mitra** has completed his B.Tech in Civil Engineering from KIIT University, Bhubaneswar in 2016 and M.Tech in Geo-Informatics from Indian Institute of Engineering Science and Technology (IEST), Shibpur in 2018. He has worked as a Guest Faculty in the Department of Civil Engineering, University Institute of Technology (UIT), Burdwan University for more than 3 years. He is currently working as an Assistant Professor in the Department of Civil Engineering, Abacus Institute of Engineering & Management, Natungram, Magra, West Bengal – 712148.



**Debaduyti Nath** has completed his B.Tech in Civil Engineering from Hooghly Engineering and Technology College (MAKAUT), W.B in 2016 and M.Tech in Geo-Informatics from Indian Institute of Engineering Science and Technology (IEST), Shibpur in 2018.