
IMAGE: A MAP OF THE STARS OF THE ORION CONSTELLATION

Print ISSN: 2631-8490 Online ISSN: 2631-8504

# JournalPreview

London Journal of Research in Science: Natural and Formal
Volume 23 | Issue 19 | Compilation 1.0

Great Britain Journals Press

WWW.JOURNALSPRESS.COM

# Journal Content
## In this Issue

Great Britain Journals Press

# Conserving a Creative living Tradition - the Living Root Bridges of North East India (Meghalaya)

*Nitin Maurya, Vivek Kumar, Augustus Suting, Wankit Swer, Genavafa Behphat & Vipin Kumar*

## ABSTRACT

Though they may not be using scientific theories or terminology, common people still practice 'science' in their 'laboratories of life'. They have a tremendous capacity to observe, assimilate and learn from the activities of day-to-day life and the environment around them. One of the many examples to prove this is the interesting case of the Living Root Bridges, which are a remarkable engineering accomplishment of the indigenous tribes of Meghalaya, a state located in the North Eastern part of India. These bridges are built using the living secondary roots of ficus trees (Ficus elastica Roxb. ex Hornem.) planted on either side of a stream or a gorge and guided across the expanse. The article highlights the efforts of the local tribal communities in pursuing this tradition and the role of stakeholders in providing an enabling environment for the sustenance and propagation of this living tradition.

*Keywords:* nature architecture, Khasi, Jaintia, ficus elastica, conservation, sustainability.

*Classification:* LCC Code: TA1-TA2040

*Language:* English

Great Britain Journals Press

LJP Copyright ID: 925681
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 23 | Issue 19 | Compilation 1.0

# Conserving a Creative living Tradition - the Living Root Bridges of North East India (Meghalaya)

Nitin Maurya[α], Vivek Kumar[σ], Augustus Suting[ρ], Wankit Swer[ↀ], Genavafa Behphat[¥] & Vipin Kumar[§]

## ABSTRACT

*Though they may not be using scientific theories or terminology, common people still practice 'science' in their 'laboratories of life'. They have a tremendous capacity to observe, assimilate and learn from the activities of day-to-day life and the environment around them. One of the many examples to prove this is the interesting case of the Living Root Bridges, which are a remarkable engineering accomplishment of the indigenous tribes of Meghalaya, a state located in the North Eastern part of India. These bridges are built using the living secondary roots of ficus trees (Ficus elastica Roxb. ex Hornem.) planted on either side of a stream or a gorge and guided across the expanse. The article highlights the efforts of the local tribal communities in pursuing this tradition and the role of stakeholders in providing an enabling environment for the sustenance and propagation of this living tradition.*

*Keywords:* nature architecture, Khasi, Jaintia, ficus elastica, conservation, sustainability.

*Author* α σ §: National Innovation Foundation – India, Gandhinagar, Gujarat, India.
ρ: State Council of Science, Technology and Environment, Shillong, East Khasi Hills, Meghalaya, India.
ρ ↀ: ¥ Meghalaya Basin Development Authority, Shillong, East Khasi Hills, Meghalaya, India. : Foundation for Empowerment and Inclusive Development, East Khasi Hills, Meghalaya.

## I. INTRODUCTION

Nature has bestowed upon humans a lot of useful resources, which they have been using since time immemorial, primarily for food, clothing and dwelling. As humans settled in agriculture, and aggregated themselves in villages, towns and cities, day-to-day dependence on forests reduced considerably. However, a lot of indigenous communities, which reside in or near forests, depend majorly on them for their sustenance. They have, over a period of centuries, developed an intimate relationship with the forests and the nature around them. This relationship is reflected in many facets of their existence, including home and architecture.

Learning from nature and incorporating natural elements in architecture is increasingly being taken up by architects and researchers. However, Vallas and Courard (2017) have argued that 'living architecture' has been in existence for centuries. Literature related to the use of natural elements to develop pathways across or over obstacles is sparse but available. A number of vine bridges have been reported from the Iya Valley, a remote mountainous valley in Tokushima Prefecture of Japan, of which three, the *Iya Kazurabashi* and the *Oku-Iya Kazurabashi* bridges, are known to be maintained to date. The wooden slabs used in these bridges are held together strongly by the climber *Wisteria floribunda* (Willd.) DC. resulting in these bridges reaching up to the length of 43m (ibid). These vine

bridges were built over deep gorges and were used to transport goods and people and could be easily cut in case of any invasion (Luck 2019). Spanning about 25m over the Batang Bayang River, the *Jembatan Akar* (Root Bridge in the Indonesian language) on the island of Sumatra in Indonesia, is knitted from the roots of two banyan trees and developed over a period of approximately 25 years beginning in around 1890 (Boscamp 2013, Grundhauser n.d.). On the Java island of Indonesia, the Baduy people are reported to have built a root bridge over the Cisimeut River, which connects them with the outside civilization (Anonymous 2019).

The Incas in the Andean region of Peru are also known to have constructed a large number of bridges along with their elaborate road network. Most of these bridges were over gorges and were made of grass and other material (Bauer 2006). Built with perishable material, not many remain today. However, the last remaining Inca rope bridge, the Quesuachka, over the Apurimac River in the Quehue district of Peru could still be seen (ibid). In India as well, the Khasi and Jaintia people of Meghalaya are known to develop root bridges over the last many centuries. A root bridge has also been recently reported from Nagaland state as well (Shachar 2016).

## II. THE LIVING ROOT BRIDGES (LRBs) OF MEGHALAYA

Meghalaya, located in North-East India, is home to the Khasi, Garo and Jaintia tribes, who together are the world's largest matrilineal societies. Predominantly a hilly state with stretches of valleys and plateaus, Meghalaya has an agrarian economy. The state receives heavy rainfall during the monsoons with Cherrapunji and Mawsynram, one of the world's two wettest places, being located in the East Khasi district of the state (Guhathakurta et al. 2020).


*Fig. 1:* Location of Meghalaya in India [Map credit: www.mapsofindia.com (India map), www.southwestgarohills.gov.in (Meghalaya map)]

The villages of Khasi and Jaintia people residing in the southern part of Meghalaya are located in the hills interspersed with waterfalls, rivers and rivulets. Located in a heavy rainfall region, it was difficult for people to cross these hilly streams, especially during the monsoons. Noticing that other solutions like wooden or bamboo bridges could not withstand the strong currents and heavy rains, the people thought of an ingenious idea. They utilised the strength, vigour and tenacity of the aerial roots of the Indian rubber fig (*Ficus elastica* Roxb. ex Hornem.) tree to build the Living Root Bridges (*jing kieng jri* in Khasi language) across the hilly terrain for the transportation of goods and people. While the LRBs have been in existence for centuries, scientific studies on them have been meagre until recently when a lot of literature has emerged. Mathew (2005) has described the LRBs while Shankar (2015) has studied and detailed the construction of the LRBs comparing them with steel suspension bridges.

Chaudhuri et. al. (2016) have described LRBs as a low-cost bio-engineering technology, which can be replicated in other similar regions as well. Ludwig et. al. (2019), while studying the LRBs from an interdisciplinary perspective, suggest taking up studies on *F. elastica* from a structural engineering perspective so as to scientifically understand the development process of LRBs, which may be helpful for future research on living architecture. Middleton et. al. (2019) describe the findings of a photogrammetric survey of the living root bridges. Middleton et. al. (2020) have also evaluated LRBs as an example of regenerative design.

The Khasis and Jaintias make use of the secondary aerial roots of the *Ficus elastica* trees to build these LRBs across streams or gorges. The aerial roots of the Ficus trees on one side of the stream are guided to the other side so that they can take root firmly in the soil. Sometimes, bamboo or wood scaffolding is provided to support the roots during this period. Over a period of many years, these roots grow, get thicker and stronger, and are able to take the weight of people (about 20 of them at a time and sometimes more) walking on them. Sometimes these bridges are reinforced with secondary support material like wood pieces and stones. Depending upon its span, the development of a sturdy living root bridge can take anything from 10-25 years and one which can last for centuries.

*Fig. 2:* Jingmaham Living Root Bridge, Mawlynnong, East Khasi Hills, Meghalaya (Photo: Nitin Maurya)

An individual takes the lead in the construction of the LRB and is supported by the community. For safeguarding the LRB, there are community restrictions on extracting latex of *Ficus* trees used for building the LRBs or using any other parts or that of other trees in the vicinity. Mining sand and stones from the area around the LRBs, especially near the roots of the tree(s), is also prohibited. The community, mobilised through the *Darbar* (village council) or the village societies/clubs, maintains the LRBs periodically checking the roots, reinforcing the bridge with wood planks or stones, or replacing them or weaving new roots into the bridge. Sometimes during the rains, strong water currents of the stream wash away the soil, sand or stones underneath the bridges. In such cases, the community members lay stones/boulders under the bridge to protect the area from water erosion. While it is difficult to ascertain the life of a LRB, community elders can identify a root getting weak (evident by its dryness). In such a case other available roots can be pulled in and weaved into the bridge to provide it the required strength.

## III. THE FICUS ELASTICA, THE KHASIS AND THE JAINTIAS

Humans are connected to and are dependent on nature in multiple ways. Many indigenous cultures demonstrate that a harmonious co-existence with nature is possible by employing social and cultural mechanisms to safeguard and responsibly use available natural resources to meet community needs. That the survival of forests/ nature is important for the existence of humans themselves, is stressed by all indigenous communities the world over. Like other traditional societies, the lives of the Khasis and the Jaintias are also closely intertwined with the forests around them. Forests for them are sacrosanct and not just a natural resource available to exploit. The LRBs, similarly, are also deeply ingrained in the culture of the communities. The process of development of the LRBs has evolved over many centuries and is part of the inherited knowledge system of the communities; knowledge that has been orally transferred over generations. Shankar (2016) deems LRBs as an outstanding example of the Khasis' sacred relation with nature.

The decision to use *F. elastica* to develop root bridges has to do both with the biological properties of the roots of the plant, and the keen and careful observation of nature and the environment by the Khasis and Jaintias. Heavy rains during the monsoon months make travelling difficult in dense forests interspersed with mountain streams and gorges. Under these conditions, while other material degrades over a period of time, the LRBs grow sturdier. A member of the Moracece family, *Ficus elastica* is a large tropical tree found in evergreen forests, particularly along river banks (CABI 2020). *F. elastica* generally grows as an epiphyte (growing on another tree) with its aerial roots growing downwards to reach the ground, becoming thicker and robust. Gradually, the aerial roots encircle the host tree, and its foliage smothers the host resulting in the death of the host tree. Zimmermann et al. (1968) in their study on the aerial roots of *F. benjamina* L. found out that post anchoring on a substrate they produce tension wood, which facilitates the inosculation (fusing and growing together) process of the aerial roots. Though similar properties of *F. elastica* have not been studied in detail, however coming from the same genus it may be assumed that the process is similar in this species. Shankar (2015) opines that the aerial roots' capacity to react to mechanical stress with secondary growth, formation of tension wood and development of inosculations, together enables them to perform the structural support function. Another observation that may have prompted the Khasis to use the *F. elastica* may be its fast-growing habit and tight soil-holding property, which may be evident by lesser landslides in the region of their growth. Shankar (2017) has mentioned the use of indigenous materials, traditional tools and of the expertise developed over many generations, in the development of these living root bridges.

## IV. CONSERVING, PRESERVING AND SUSTAINING THE LIVING TRADITION: THE LRBS

Gupta (1999) has argued that conservation and preservation have to be undertaken at both resource and associated knowledge levels. He opines that if only resources are conserved, knowledge will get eroded over time. If only knowledge is conserved, erosion of resources will become an eventuality. For the sustainability of the system, as a whole, incorporating indigenous/ local knowledge in the strategies of regeneration of the resources itself thus becomes important (ibid). Shankar (2017) has opined that the Living Root Bridge is a threatened community-based but sustainable infrastructure practice. He suggests that conservation efforts, research and grassroots entrepreneurship together can possibly provide a viable solution for reinvigorating and contemporising traditional knowledge practices. Gupta (2001) remarks that many communities have respect for nature and its conservation, and concern for future generations. However, incentives for conservation efforts are necessary (Tamale 1996; United Nations Environment Programme 1996, 1998; Emerton 2000).

Appreciating the creativity and unacknowledged scientific acumen of the Khasi and Jaintia tribes, the National Innovation Foundation - India (NIF), an autonomous body of the Indian Government's Science and Technology department, recognised this unique community-developed, community-nurtured grassroots innovation in its Ninth National Biennial Awards (National Innovation Foundation – India 2017). The award money (approx. $6800) was to be used by the community(ies) for any community development-related work only. Many rounds of meetings and discussions were held among the village councils (*darbars*), community members, Meghalaya Basin Development Authority (MBDA) and State Council of Science, Technology and Environment (SCSTE), both constituents of the Government of Meghalaya in consultation with the NIF. A number of ideas were discussed but a consensus could not be reached. While the discussions were ongoing, to earn interest on the amount, the award money was parked with the SCSTE.

The challenge was to utilise the award money and not simply use it. During the ensuing discussions, first a concept for building a 'Living Root Bridge Museum' came up and then from it the concept of a 'Living Root Bridge Museum and School (LRBMS)' emerged. The idea is to involve the community toset up a Museum beside a young (few years old) living roots bridge where pictures and videos of the 'growing' root bridge could be added periodically along with other related items. Thus, the Museum 'lives and grows' along with the living roots bridge. To involve community youth in understanding the process of development of the root bridges, explaining to them the science behind it and undertaking research, a 'School' has also been proposed. The site is proposed to be community-built, community-managed and open to tourists not only to enjoythe beauty of the place but also to understand and appreciate the ingenuity of tribal communities who have lived in harmony with nature for centuries. The LRBMS will also be expected to facilitate scientific research on the LRBs by institutions/ researchers desirable for the same. Many associated activities like maintaining sanitation, providing drinking water facilities, maintenance of the earmarked area and buffer zone, sale of local food and crops to visitors, etc. are also expected to come up around the LRBMS, resulting in more livelihood opportunities for the local people.

The setting up of the 'Living Root Bridge Museum and School', thus, is expected to address three core issues

a. *Conserving* - the Living Root Bridges.
b. *Preserving* - the associated knowledge system and institutionalising its transfer to the next generation.
c. *Sustaining* - through community-level entrepreneurship in managing the museum, school and ancillary activities.

While appreciating the concept, the Government of Meghalaya topped up the award money with a corpus of approximately $34000 for the purpose. The SCSTE has been undertaking elaborate fieldwork and consultation with the communities to identify a suitable site for developing the Living Root Bridges Museum and School. Narrowing down on the Pynursla Community & Rural Development block of East Khasi Hills district of Meghalaya, the SCSTE has been able to identify and document 71 living root bridges in 13 villages from the block with their span ranging from 2- 52 meters and age 35-300 years. This work is being taken up cautiously taking all the stakeholders in confidence, especially the local tribal community. The MBDA has also been making considerable efforts to document and encourage the conservation and preservation of the living roots bridges across the state.

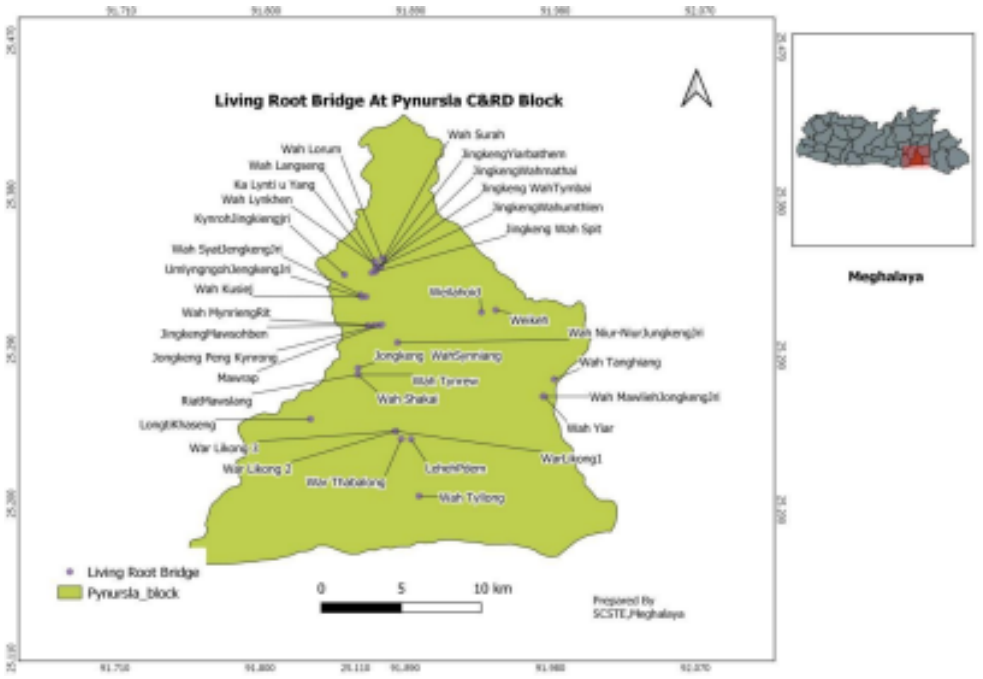


*Fig. 3:* Location of living root bridges in Pynursla C&RD block, East Khasi Hills, Meghalaya (Source: SCSTE, Meghalaya)

Once the site for developing the LRBMS is identified in partnership with the local communities, the work to set it up will be initiated. It is hoped that this approach of conservation linked with grassroots-level entrepreneurship will facilitate the continuity of this living tradition over the next many centuries. Though the future will evaluate the success of this initiative, this definitely has the potential to present a case for sustainable conservation and preservation.

## ACKNOWLEDGEMENTS

The authors acknowledge the generous support of their respective institutions, team members, Khasi community and in particular the Headmen and community members of the villages under the Pynursla Community & Rural Development Block, East Khasi Hills, Meghalaya.

## REFERENCES CITED

1. Anonymous (2019) The Baduy Root Bridge is Still Saves a Mystery. https://travellingto.asia/the baduy-root-bridge-is-still-saves-a-mystery/. Accessed 16 April 2023.
2. Bauer BS (2006) Suspension Bridges of the Inca Empire. In: Isbell WH, Silverman H (eds) Andean Archaeology III, Springer, Boston, MA, pp 468-493 https://doi.org/ 10.1007/0-387-28940- 2_19.
3. Boscamp E (2013) Jembatan Akar: Indonesia's Living Root Bridge. Huffpost. https://www.huffingtonpost.in/entry/living-root-bridge-indonesia_n_3468033. Accessed 15 March 2023.
4. CABI (2020) Invasive Species Compendium, Wallingford, UK. https://www.cabi.org/isc/ datasheet/24090. Accessed 15 May 2023.
5. Chaudhuri P, Bhattacharyya S, Samal AC (2016) Living Root Bridge: A potential no-cost eco-technology for mitigating rural communication problems. Int J Exp Res Rev 5:33-36.
6. Emerton L (2000) Using Economic Incentives for Biodiversity Conservation. IUCN — The World Conservation Union. https://portals.iucn.org/library/efiles/documents/PDF-2000-002.pdf. Accessed 26 March 2023.
7. Grundhauser E (n.d.) Jembatan Akar- West Sumatra, Indonesia. https://www.atlasobscura.com /places/jembatan-akar. Accessed 16 April 2023.

8. Guhathakurta P, Bhagwat PP, Satpute US, Menon P, Prasad AK, Sable ST, Advani SC (2020) Observed Rainfall Variability and Changes Over Meghalaya State. In: Met Monograph No.: ESSO/IMD/HS/Rainfall Variability/17(2020)/41, India Meteorological Department, Pune.
9. Gupta AK (1999). Compensating local communities for conserving biodiversity: How much, who will, how and when. Researchgate. https://www.researchgate.net/publication/46437138_Compensating_Local_Communities_for_Conserving_Biodiversity_How_Much_Who_Will_How_and_When. Accessed 12 May 2023.
10. Gupta AK (2001) How Can Asian Countries Protect Traditional Knowledge, Farmers Rights and Access to Genetic Resources through the Implementation or Review of the WTO TRIPS Agreement?http://anilg.sristi.org/wp-content/Papers/How%20Can%20Asian%20 Countries %20Protect%20Traditional%20Knowledge.doc. Accessed 18 May 2023.
11. Luck A (2019) 4 reasons to visit Japan's remote Iya Valley on Shikoku Island. Tokyo Weekender. https://www.tokyoweekender.com/2019/04/4-reasons-to-visit-japans-remote-iya-valley on-shikoku -island/ . Accessed 16 May 2023.
12. Ludwig F, Middleton W, Gallenmüller F, Rogers P, Speck T (2019) Living bridges using aerial roots of ficus elastica – an interdisciplinary perspective. Sci Rep 9:12226. https://doi.org/10.1038/s41598-019-48652-w.
13. Mathew R (2005) The living roots bridges of Meghalaya. Current Science 89(1):10- 11.
14. Middleton W, Shu Q, Ludwig F (2019) Photogrammetry as a tool for living architecture. Int Arch Photogramm Remote Sens Spatial Inf Sci 42:195–201.
15. Middleton W, Habibi A, Shankar S, Ludwig F (2020) Characterizing regenerative aspects of living root brishankardges. Sustainability 12,3267 doi:10.3390/su12083267.
16. National Innovation Foundation – India (2017) 9th National Biennial Grassroots Innovation and Outstanding Traditional Knowledge Awards. http://nif.org.in/dwn_files/9th-Award Book.pdf. Accessed 16 April 2023.
17. Shachar G (2016) Living Root Bridges of Nagaland India – Nyahnyu Village Mon District. https://guyshachar.com/en/2016/living-root-bridges-nagaland-india-mon-myahnyu/. Accessed 15 May 2023.
18. Shankar S (2015) Living root Bridges: State of knowledge, fundamental research and future application. IABSE Symposium Report 105(4):1-8 doi:10.2749/222137815818359474.
19. Shankar S (2016) Revitalizing traditional knowledge: Living Root Bridge as a Biome. http://www.sanjeevshankar.com/pdf/Living-Root-Bridge-as-a-Biome-Sanjeev-Shankar SUSCON -5-IIM-Shillong-2016.pdf. Accessed 15 May 2023.
20. Shankar S (2017) Revitalizing Plant-based Knowledge in Northeast India. The Solutions Journal 8(2). https://www.thesolutionsjournal.com/article/revitalizing-plant-based-knowledge north-east- India/ Accessed 15 November 2022.
21. Tamale ES (1996) Incentive measures for the conservation and sustainable use of biological diversity in Uganda: a case study of the 'development through conservation' project in communities around Bwindi National Park. Paper presented at the Workshop on Incentives for Biodiversity: Sharing Experiences, Montreal, Canada, 30 August - 1 September 1996 https://www.cbd.int/doc/case-studies/inc/cs-inc-ca-02-en.pdf.
22. United Nations Environment Programme (1996) Sharing of Experiences on Incentive Measures for Conservation and Sustainable Use. In: Note by the Executive Secretary. Convention on Biological Diversity, UNEP Document No. UNEP/CBD/COP/3/24, 20 September, 1996. Accessed 26 April 2023.
23. United Nations Environment Programme (1998) Design and Implementation of Incentive Measures. In: Note by the Executive Secretary. Convention on Biological Diversity, UNEP Document No. UNEP/CBD/COP/4/18, 1 February, 1998. Accessed 26 April 2023.
24. Vallas T, Courard L (2017) Using nature in architecture: Building a living house with mycelium and trees. Frontiers of Architectural Research 6(3): 318-328. https://doi.org/10.1016/j.foar.2017.05.003.s
25. Zimmerman MH, Wardrop AB, Tomlinson PB (1968) Tension wood in aerial roots of Ficus benjamina L. Wood Science and Technology 2: 95-104.

*This page is intentionally left blank*

# 3-D Seismic Attribute Analysis and Modeling for Reservoir Pore-Fill Interpretation in an X-Field of the Niger Delta, Southern Nigeria

*Ogini Arthur, A , Alile, O .M. & Ugbena Kelvins, G*

*Federal University Otuoke*

## ABSTRACT

3-D seismic attribute analysis and reservoir pore-fill characterization were conducted in an X-field in the Niger Delta petroleum province Nigeria. The main objective of the study was to look for possible relationships between selected seismic attributes and reservoir fluid types in the X-field with a view to erecting a model for reservoir pore-fill interpretation based on seismic attribute analysis in the X-field. To achieve the expected results, the following data sets were used: 3-D seismic sections and maps (amplitude, frequency, phase and polarity display/maps) of a selected horizon pseudo-named (M10). Also compiled were digital well logs which were combined in an integrated analysis to arrive at the final results. It was observed that seismic reflection polarity values above 4,574 were indicative of good reservoir sands. In contrast, values less than -3,520 were seen to represent shale with good sealing capacity. Amplitudes less than 13,480 signified an unfavourable event while values between 13,840 and 17,789 were indicative of oil, whereas all true amplitude greater than 19,764 represented gas. Frequencies between 18.6 – 17.9 Hz were seen to represent oil while values below 17.0 Hz corresponded to gas. Finally, the reservoir pore-fills in the chosen horizon (M10) were interpreted to be gas; around well (7) and the western limits of the horizon, and oil; around well (4). In conclusion, a pre-drilling guide or model for reservoir pore-fill interpretation from seismic attribute analysis of 3-D seismic data volume in the X -field was specified.

*Keywords:* well-log controlled seismic attribute modeling, reservoir pore-fill interpretation.

*Classification:* LCC Code: TN260-TN263

*Language:* English

Great Britain Journals Press

LJP Copyright ID: 925682
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 23 | Issue 19 | Compilation 1.0

# 3-D Seismic Attribute Analysis and Modeling for Reservoir Pore-Fill Interpretation in an X-Field of the Niger Delta, Southern Nigeria

Ogini Arthur, A[1] , Alile, O .M.[2] & Ugbena Kelvins, G.[3]

## ABSTRACT

*3-D seismic attribute analysis and reservoir pore-fill characterization were conducted in an X-field in the Niger Delta petroleum province Nigeria. The main objective of the study was to look for possible relationships between selected seismic attributes and reservoir fluid types in the X-field with a view to erecting a model for reservoir pore-fill interpretation based on seismic attribute analysis in the X-field. To achieve the expected results, the following data sets were used: 3-D seismic sections and maps (amplitude, frequency, phase and polarity display/maps) of a selected horizon pseudo-named (M10). Also compiled were digital well logs which were combined in an integrated analysis to arrive at the final results. It was observed that seismic reflection polarity values above 4,574 were indicative of good reservoir sands. In contrast, values less than -3,520 were seen to represent shale with good sealing capacity. Amplitudes less than 13,480 signified an unfavourable event while values between 13,840 and 17,789 were indicative of oil, whereas all true amplitude greater than 19,764 represented gas. Frequencies between 18.6 – 17.9 Hz were seen to represent oil while values below 17.0 Hz corresponded to gas. Finally, the reservoir pore-fills in the chosen horizon (M10) were interpreted to be gas; around well (7) and the western limits of the horizon, and oil; around well (4). In conclusion, a pre-drilling guide or model for reservoir pore-fill interpretation from seismic attribute analysis of 3-D seismic data volume in the X -field was specified .*

*Keywords:* well-log controlled seismic attribute modeling, reservoir pore-fill interpretation.

*Corresponding Author:* Department of Physics/Geology, Federal University Otuoke, Bayelsa State, Nigeria.
2 Department of Physics, University of Benin, Edo state, Nigeria.
3 Department of earth science, Kogi State University, Anyigba-Nigeria.



## I. INTRODUCTION

According to Kearey and Brooks (1991), "the ability to identify sedimentary environments and predict litho facies from analysis of seismic sections can be of great value to exploration programs, providing a pointer to potential source rocks". The above quote is of exploration significance in locating reservoir and seal rocks, but for the purpose of estimating the hydrocarbon in place and perhaps carrying out production optimization plans; not only will the reservoir engineer be interested in variables such as the reservoir structural/stratigraphic frameworks, petrophysical properties, temperature/pressure regimes but of major concern to him is the fluid type (oil and gas) in the pore spaces of the reservoir where production is intended. Looking at the reservoir without knowledge of the pore-fills is like viewing a container from the outside without knowledge of what is inside. Because wildcat drilling is capital intensive, a number of methods have been adopted with a view of providing a solution to the above problem, one of which is the 'bright spot' technology which allows hydrocarbon to be inferred directly from 'true amplitude seismic sections' by the use of Direct Hydrocarbon Indicators (DHI). These high amplitude events are attributable to the large reflection coefficient at the top and bottom of gas zones (typically gas filled sands) within a hydrocarbon reservoir (Sheriff, 1975; Sheriff, 1978). Since

well-logs, which contain a whole lot of information, are considered one dimensional in comparison to the areal extent of a typical hydrocarbon field; 3-D seismic on the other hand provides information about the second and third dimensions of the subsurface lithology, structures and the associated fluids in a field (Sheriff & Geldart, 1995). Much valuable information is contained in the 'True' amplitude seismic section; any lateral change in reflection amplitude is due to lateral changes in the lithology of the rock unit or its fluid content. The same can be said of the other attributes (Meckel & Nath, 1977; Sheriff, 1980; Chao et al., 2009; Sheriff & Geldart 1995). Seismic attributes are characteristic parameters extracted from seismic traces through given mathematical methods, they have the ability to reflect intrinsic characteristic of seismic information from various angles depending on the type of attributes and the aim/manner in which they are computed (Chao, et al., 2009; Yao & Chopra, 2000). Seismic attribute studies have a theoretical foundation from complex trace analysis commonly used in electrical engineering and signal processing (Gabor, 1946; Bracewell, 1978; Cramer et al., 1967; Oppenheim & Schafer, 1975; Robertson et., al, 1988). In seismic data processing and interpretation; the term complex seismic trace was used to describe a seismic wave function made of a 'Real' and an 'Imaginary or complex' parts. The seismic trace from the surface seismic survey is treated as the 'real' part of the complex trace, while the 'imaginary' part is the Hilbert transform of the 'real' part. Such a representation allows for an easy and natural way of separating the seismic reflection amplitude (length of phasor) from the reflection 'angle' (which defines frequency and phase) making possible the computation of instantaneous attributes (Tarner et al., 1979; Barns, 1991; Bodine,1984; Brown, 1996). "When evaluated and sampled using appropriate time windows, the envelope of instantaneous amplitude combined with instantaneous frequency and phase improves resolution, hence seismic events from reservoirs are more clearly defined" ( Zang & Bentley, 2008). The following are some recent work in the study area: Obiekezie and Omoja (2019) used seismic attribute analysis to delineate hydrocarbon prospective zones in parts of the on-shore Niger delta "Uzot-Field" where the aim was to identify faults, seals and roll-over anticlines; Root Mean Square (RMS) amplitude was employed in the identification of targets. Also, reservoir prediction and prospect identification studies were carried out using seismic attribute analysis in 'OK-Field' of the Niger Delta. Instantaneous amplitude aided identification of fluid contacts by the application of 'bright' and 'dim' spots technology (Ologe & Olowokere, 2021). Again, Opara and Osaki (2018) exemplified the viability of integrating structural interpretations, petrophysical evaluation and seismic attribute analysis in evaluating the hydrocarbon potential of 'Opu-Field' in the coastal swamp Niger Delta depobelt. The major attributes considered were root mean square amplitude and instantaneous frequency. Augustine et., al (2021) has combined structural/stratigraphic interpretations and seismic attribute analysis to assess the sealing capacity of growth faults in a section of the Niger delta. Though the 'variance' attribute was used to infer structural zones from the seismic section, it was not directly employed as a quantitative fluid interpretation parameter. Also, in the Baris Oil and Gas field of the Niger delta, well log derived estimates of porosity, permeability and fluid saturation from subsurface data were used for the petrophysical evaluation of reservoirs on one hand. On the other hand, information about the reservoir geometry, trends and extent of associated listric faults were derived from seismic data (Bate et al., 2023). Furthermore, in the reservoir characterization of the Otan-Ile field section of the Niger delta; qualitative use of seismic amplitude attribute was used to reconfirmed the presence of hydrocarbon by superimposing the former on a depth structure map (Tokunbo et., al , 2021). But the current study is focused on the use of seismic attribute analysis to discriminate among reservoir fluids (oil, water and gas) both qualitatively and quantitatively before drilling.

## II. THEORETICAL BACKGROUND

The phasor representation of the complex seismic trace is given by the equation:

$$C(t) = A(t)\{sin[\Theta(t)]\} \quad 1$$

Where
$C(t)$: the complex seismic trace function
$A(t)$: the amplitude of the complex seismic trace (length of phasor)
$\Theta(t)$: Stands for frequency and phase (angle of phasor).

From equation (1) above, the complex seismic trace is made up of real and imaginary parts; the real part is the actual surface seismic recorded signals obtained from geophones, and is given by:

$$f(t) = A(t)cos[\Theta(t)] \quad 2$$

Where $A(t)$ and $\Theta(t)$ are the respective functions of the amplitude and phase of the complex seismic trace.

The complex seismic trace, otherwise called the 'quadrature' of the 'real trace' is the 90° rotation of the real trace calculated using the Hilbert transform in a 3-D plot (Bentley et al., 1999; Bracewell, 1978; Taner et al,. 1979)

Equation (2) is realizable from equation (1) by digitization using appropriate time window, and its Hilbert transform is derived according to the following steps:

1. 90° phase shifting of the real component f(t)
2. performing an inverse Fourier transform on f(t)

The result from this process is the negative of the Hilbert transform and equation (1) can now be written as:

$$C(t) = f(t) + ih(t) \quad 3$$

Where $ih(t)$ represents the Hilbert transform of $f(t)$ and

$C(t)$: a helical function that spirals about the time axis.

The amplitude or 'true amplitude' is the maximum absolute amplitude for all 'phase' rotations at a given time sample, it corresponds to the envelope of the complex seismic trace mathematically given by;

$$/C(t)/ = \sqrt{\{f(t)^2 + h(t)^2\}} \quad 4$$

Where;
$/C(t)/$: is the true amplitude
$f(t)$ and $h(t)$: $are$ the amplitudes of the real and imaginary traces of $C(t)$ at time 't' respectively.

The instantaneous phase is the time-by-time sample phase value of the complex seismic trace envelope given by;

$$\theta = [h(t)/f(t)] \quad 5$$

$\theta$: is the instantaneous phase;
$f(t)$, $h(t)$: are the amplitudes of the real and imaginary traces

We define instantaneous frequency as the time derivative of instantaneous phase or a measure of how quickly or slowly phase changes as a function of time, which is not to be confused with Fourier transform frequencies (the bandwidth of the inherent wave form).
Therefore;

When phase is given by $\theta(t)$, representing a combination of angle and frequency, then the frequency is found by evaluating the time derivative of $\theta(t)$.

$$f = \frac{1}{2\pi} d[\theta(t)]/dt \tag{6}$$

$$\frac{d[\theta(t)]}{dt} = 1/2\pi \frac{d[\omega t+\phi]}{dt}$$

$$= \frac{\omega}{2\pi} = f \tag{7}$$

$\omega$ represents angular frequency while $\phi$ is a phase determining constant.
Equation (7) is the normal definition of frequency.

Polarity on the other hand, is identified as the sign (whether positive or negative) of the seismic trace amplitude which is determined by the directional variation of acoustic impedance in sedimentary strata.

## III. MATERIAL AND METHOD

The following data sets were used:

- A 3-D polarity display seismic section of an X- field in the Niger Delta oil province complimented with wire-line log data from three Wells.
- Point shot (T-Z) data for the 3 Wells.
- Base maps.
- Digital well-logs comprising of gamma-ray, resistivity, sonic and density logs enhanced with porosity and water/hydrocarbon saturations information
- Seismic reflection amplitude, frequency and phase data for the chosen horizon.
- Attribute maps and sections (amplitude, frequency and phase).
- Reflection time data for the chosen horizon pseudo named M10.

The method adopted involved the use of in-lines and cross-lines (bines and stacks) to map the horizon of interest across the field, and posting reflection times at the top and bottom of the reservoir (horizon M10) to their corresponding seismic grids. Using the t-z function, time-depth conversions were done and the resulting depth tied to bore-hole data at the various Wells. On the base map/ seismic reflection data, the positions of the 3 Wells were identified and a vertical slice/section was cut through the wells to enhance display of field data at the various Well locations on a single time axis. The four seismic attributes considered in this study were amplitude, frequency, phase and polarity. A quantitative and qualitative analysis of each seismic attribute was done in the presence of the associated reservoir fluid to reveal its responses/changes as influenced by the three different reservoir fluids (oil, water or gas). The observed seismic attribute responses at the horizon of interest (M10) were displayed as tables and maps.

## IV. RESULTS

Interpretation of abbreviations as used in the result tables:

CALP: Calliper log
FDC: Compensated density log
RT: Resistivity log
GR: Gamma ray log
POR: Porosity
SH: Hydrocarbon saturation
SW: Water saturation

*Table 1:* Digital Well-Log Extracts and Fluid Interpretation of Horizon M10 Well-7

| Depth (ft) | CALP | FDC (g/cc) | RT (Ωm) | GR (API) | POR | SH | SW | REMARKS |
|---|---|---|---|---|---|---|---|---|
| 5,174.5 | 11.91 | 2.250 | 1.976 | 85.20 | 0.207 | 0.154 | 0.947 | shaly sst/HC |
| 5,176.5 | 11.91 | 2.210 | 1.930 | 92.39 | 0.194 | 0.118 | 0.882 | Shaly sst/HC |
| 5,180.5 | 11.56 | 2.290 | 2.506 | 89.99 | 0.199 | 0.180 | 0.820 | Shaly sst/HC |
| 5,182 | 11.77 | 2.250 | 2.146 | 91.08 | 0.197 | 0.188 | 0.812 | Shaly sst/HC |
| 5,184 | 11.76 | 2.250 | 2.159 | 86.08 | 0.206 | 0.205 | 0.795 | Shaly sst/HC |
| 5,186.5 | 11.72 | 2.240 | 2.172 | 87.72 | 0.203 | 0.204 | 0.796 | Shaly sst/HC |
| 5,190 | 11.96 | 2.270 | 2.028 | 90.27 | 0.198 | 0.156 | 0.844 | Shaly sst/HC |
| 5,192.5 | 11.62 | 2.340 | 2.834 | 74.94 | 0.225 | 0.364 | 0.636 | Shaly sst/HC |
| 5,194 | 11.12 | 2.090 | 5.089 | 50.757 | 0.268 | 0.570 | 0.430 | ? |
| 5,195.5 | 12.28 | 2.090 | 27.280 | 29.262 | 0.307 | 0.807 | 0.193 | ? |
| 5,197.5 | 12.26 | 2.110 | 153.66 | 27.57 | 0.310 | 0.926 | 0.074 | sst/GAS |
| 5,230.5 | 12.54 | 2.160 | 286.52 | 30.95 | 0.304 | 0.963 | 0.037 | sst/GAS |
| 5,240.5 | 14.18 | 2.080 | 105.12 | 31.49 | 0.302 | 0.927 | 0.073 | sst/GAS |
| 5,247 | 16.98 | 1.900 | 499.66 | 22.32 | 0.319 | 0.964 | 0.036 | sst/GAS |
| 5,260.5 | 15.86 | 1.930 | 980.48 | 26.37 | 0.312 | 0.974 | 0.026 | sst/GAS |
| 5,263.5 | 11.05 | 2.090 | 1959.6 | 23.29 | 0.318 | 0.983 | 0.017 | sst/GAS |
| 5,270 | 10.67 | 2.080 | 2276.8 | 19.37 | 0.324 | 0.986 | 0.014 | sst/GAS |
| 5,329.5 | 11.87 | 2.290 | 1.589 | 88.22 | 0.200 | 0.000 | 1.000 | Shale/water |

*Note: In the above interpretation; "shale/water" is used to denote a water-filled porosity while "shaly sst/HC" stands for shaly sandstone having porosity filled with both hydrocarbon and water. ?? Stands for undefined fluid interpretation*

*Table 2:* Digital Well-Log and Fluid Interpretation for Horizon M10 Well-4

| DEPTH (ft) | SONIC (Δt) | CAL P | G.R (API) | SH | POR | SW | BLK.D g/cc | RT (Ωm) | REMARK |
|---|---|---|---|---|---|---|---|---|---|
| 5,099 | 136 | 11.6 | 100 | 0.000 | 0.21 | 1.000 | 2.398 | 1.439 | Shale/w |
| 5,100 | 130 | 11.0 | 111 | 0.000 | 0.15 | 1.000 | 2.449 | 1.378 | Shale/w |
| 5,109.5 | 111 | 10.1 | 42 | 0.000 | 0.26 | 1.000 | 2.149 | 1.628 | Shale/w |
| 5,113 | 119.2 | 9.2 | 38 | 0.658 | 0.24 | 0.342 | 2.132 | 10.56 | Oil |
| 5,116.0 | 119.9 | 9.2 | 37 | 0.908 | 0.23 | 0.092 | 2.129 | 49.89 | Oil |
| 5,129.5 | 120.8 | 9.2 | 39 | 0.954 | 0.22 | 0.046 | 2.136 | 93.82 | Oil |
| 5,130 | 121.2 | 9.0 | 44 | 0.956 | 0.29 | 0.044 | 2.161 | 305.8 | Oil |
| 5,162 | 120.9 | 9.0 | 42 | 0.905 | 0.26 | 0.095 | 2.152 | 121.5 | Oil |
| 5,166.5 | 120.9 | 9.0 | 42 | 0.855 | 0.25 | 0.145 | 2.152 | 62.88 | Oil |
| 5,173 | 124.5 | 9.1 | 44 | 0.922 | 0.23 | 0.078 | 2.161 | 64.80 | Oil |
| 5,180 | 124.9 | 9.2 | 48 | 0.987 | 0.24 | 0.013 | 2.177 | 98.6 | Oil |
| 5,190 | 124.4 | 9.2 | 47 | 0.897 | 0.25 | 0.013 | 2.174 | 90.8 | Oil |
| 5,197.5 | 121.2 | 9.2 | 49 | 0.907 | 0.24 | 0.093 | 2.197 | 77.11 | Oil |
| 5,227.5 | 115.7 | 9.2 | 57 | 0.673 | 0.21 | 0.327 | 2.216 | 9.73 | H.C silt |
| 5,390 | 115.7 | 9.5 | 87 | 0.694 | 0.15 | 0.306 | 2.345 | 9.77 | H.C silt |

*Table 3:* Summarized Attribute Logs for Wells-7 and 4

| | WELL-7 | | | | WELL-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | Amplt. | Freq(Hz) | Phas/amplt | | Depth(ft) | Amplt. | Freq(Hz) | Phas/amplt. |
| | 19,764 | 15.7 | 0.19 | | 1,540 | 15,841 | 17.8 | −0.07 |
| | 19,764 | 15.7 | 0.21 | | 1,550 | 17,789 | 23.8 | −0.07 |
| | 20,751 | 16.5 | 0.22 | | 1,560 | 17,789 | 17.9 | −0.09 |
| | 22,726 | 16.8 | 0.21 | | 1,570 | 16,802 | 18.1 | − 0.07 |
| | 19,764 | 17.0 | 0.20 | | 1,580 | 14,827 | 18.9 | −0.05 |
| | 11,865 | 23.8 | 0.07 | | 1,590 | 13,840 | 18.6 | −0.07 |
| | | | | | 1,600 | - | 23.0 | −0.34 |
| | | | | | 1,620 | 7,916 | | −0.34 |

*Table 4:* Seismic Attributes and Fluid Interpretation Correlation Table for Well-7

| DEPTH (ft) | DEPTH (m) | SEISMIC ATTRIBUTES | | | FLUID TYPE |
|---|---|---|---|---|---|
| | | AMPLT | FREQ | PHASE/AMPLT | |
| 5,164.5 | 1,570 | 19,764 | 15.7 | 0.19 | GAS |

| | | | | | |
|---|---|---|---|---|---|
| 5,197.0 | 1,580 | 19,764 | 15.7 | 0.21 | GAS |
| 5,230.0 | 1,590 | 20,751 | 16.5 | 0.22 | GAS |
| 5,247.0 | 1,595 | 22,726 | 16.8 | 0.21 | GAS |
| 5,263.0 | 1,600 | 19,764 | 17.0 | 0.20 | GAS |
| 5,329.0 | 1,620 | 11,865 | 23.8 | 0.77 | UN |

*Table 5:* Seismic Attributes and Fluid Interpretation Table for Well-4

| DEPTH (ft) | DEPTH (m) | SEISMIC ATTRIBUTES | | | FLUID TYPE |
|---|---|---|---|---|---|
| | | AMPLT | FREQ | PHASE/AMPLT. | |
| 5,066 | 1,540 | 15,814 | 17.8 | 0.07 | – |
| 5,098 | 1,550 | 17,789 | 23.8 | −0.06 | UN |
| 5,131.5 | 1,560 | 17,789 | 17.9 | −0.09 | OIL |
| 5,164.5 | 1,570 | 16,802 | 18.1 | −0.07 | OIL |
| 5,197.0 | 1,580 | 14,827 | 18.9 | −0.05 | OIL |
| 5,230 | 1,590 | 13,840 | 18.6 | – 0.07 | OIL |
| 5,263 | 1,600 | – | 23.0 | −0.34 | UN |
| 5,329 | 1,620 | 7,916 | – | −0.34 | UN |

## V. DISCUSSION

Analysis of the seismic attributes/ Well data tie (Table 3) shows that; in the vicinity of Well-7, 'true amplitude' values of 19,764 and 20,751 were observed at depths 5,164.5 ft (1,570 m) and 5,230 ft (1,590 m) respectively. The well-log interpretations confirmed this depth range of the reservoir to have a gas-filled porosity (Table 1). At depths 5,247 ft (1,595 m) and 5,263 ft (1,600 m), the observed amplitudes were 22,726 and 19,764 respectively; this depth range has also been interpreted to have a gas-filled porosity (Table 4). At a depth of 5,329 ft (1,620 m), an amplitude value of 11,865 was observed, but the well log interpretation sees this depth as unfavorable (neither oil nor gas). Frequency values of 15.7 Hz, 16.5 Hz, 16.8 Hz and 17.0 Hz were observed at depths of 5,197 ft (1,580 m), 5,230 ft (1,590 m), 5,247 ft (1,595 m) and 5,263 ft (1,600m) respectively; the quoted depth ranges yet again fell into a gas zone according to the well-log interpretations. Within the same depth range, phase amplitude values of between 0.9 and 0.22 were observed.

At Well-4, slightly different results were observed; a depth range of about 5,066 ft (1,540 m) to 5,098 ft (1,550 m) was interpreted as unfavorable (Table 2) with observed 'true amplitude' readings between 15,814 and 17,789 (Tables 3 and 5). The corresponding frequency values were found to be 17.8Hz and 23.8Hz respectively. Between depths 5,151.5 ft and 5,230 ft; the 'true amplitude' attribute values fell

between 17,789 and 13,840 and the corresponding digitized frequency values were between (17.9 and 18.6) Hz. The reservoir pore-fill in this depth range was interpreted to be oil, whereas a 'true' seismic amplitude value of 7,916 was registered at depth 1,620 m, which was correlated to be unfavorable (Tables 3 and 5). A quantitative/qualitative comparison of the gamma-ray log/seismic attribute tie showed that events with very high positive polarity coincided with low gamma ray signatures, depicting sand bodies (Table 1, 2 and 3). The positivity of the polarity is a function of the cleanness of the sand body, which was found to be in agreement with the well-log derived porosity estimates. On the other hand, negative polarities were seen to be associated with shale, the negativity of the polarity values indicated the degree of 'shaliness', and all high negative polarities were seen to be associated with good sealing shale at the top and bottom of reservoirs (Tables 1 and 3). Polarity values between (4,575– 11,263) have been interpreted to represent sandstone while the value for shale stands between (-3,520 — -11,264). Polarity values between the ranges quoted above were interpreted to be indicative of lithologic boundaries and fault zones. Figure (1) is a structural map showing the posted reservoir pore-fill interpretations derived from the integrated seismic attributes and well log analysis; it could be seen that Well-7 is predominantly gas and Well-4 was found to be associated with oil, while Well-21 was interpreted to be unfavorable.

## VI. CONCLUSION

It was concluded that the reservoir pore-fill of the X-field, with particular reference to horizon M10, could be interpreted from 3-D seismic data according to the model presented in Table (6), the fluid interpretation is as shown in figure (1).

*Table 6:* 3-D Seismic Attribute Reservoir Pore-fill Interpretation Model for the X-Field

| SEISMIC ATTRIBUTES | RANGE OF VALUES | INFERENCE |
|---|---|---|
| Polarity | 4,575 – 11,263<br>-3,520 – -11,264 | Sand stone<br>shale |
| True amplitude | >19,764<br>13,840 – 17,789<br><13,840 | Gas<br>Oil<br>unfavorable |
| Frequency | >20.0<br>18.6 – 17.9<br><17.0 | Unfavorable<br>Oil<br>gas |
| Phase amplitude | continuity | continuity |

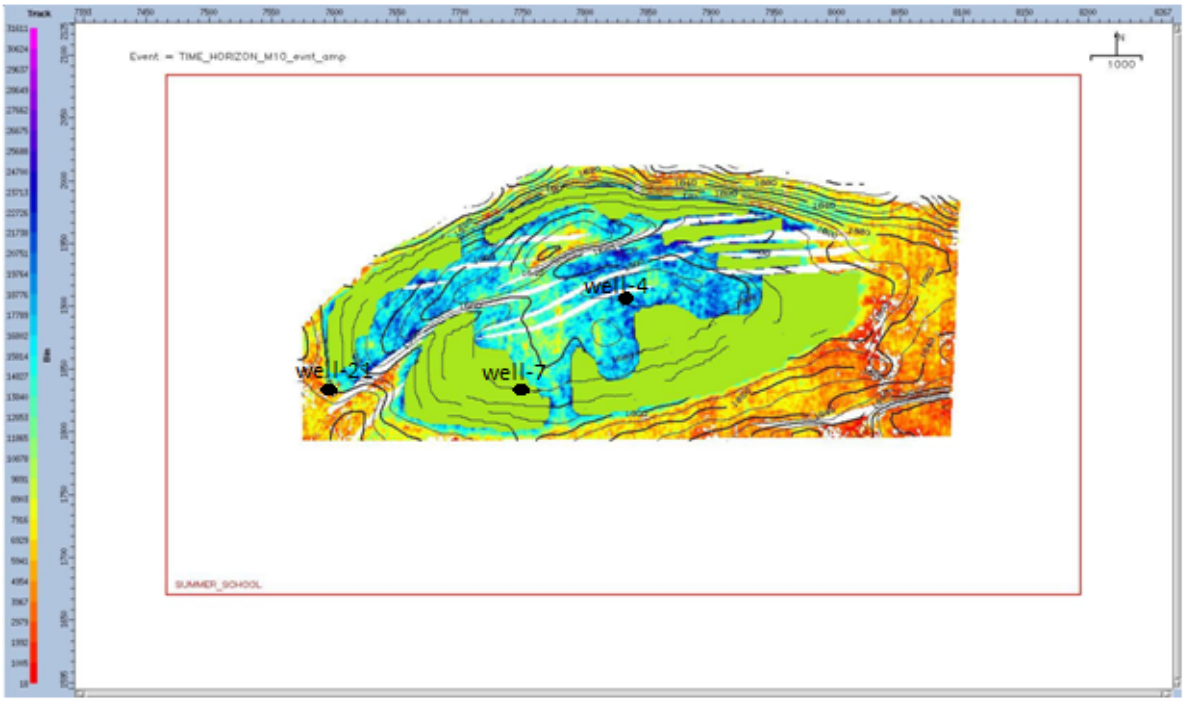


*Figure 1:* Structural Map Showing Integrated Well Log/Seismic Attribute Reservoir Pore-fill Interpretation of Horizon M10

**Legend**

| Gas | Oil | Unfavourable | unfavorable |
|---|---|---|---|

## REFERENCES


1. Augustine, O. O., Emmanuel, A. K., Clement, O. U (2021). Review of Opkara et., al "3-Dimensional seismic interpretation and fault seal assessment of Ganga Field, Niger Delta, Nigeria", journal of Environmental geology; 5(5), pp 1-8.
2. Barns, Arthur E., (1991). Instantaneous frequency and amplitude at the envelope peak of a constant phase wavelet: Geophysics, v.56, p.1058-1060.
3. Bate, B. B., Boboye, A. O., Fozao, K. F., Ndip, E. A., and Anene, N. O (2023). Petrophysical Characterization and 3-D Seismic Interpretation of Reservoirs in the Baris Field, Onshore Niger Delta Basin, Nigeria, Energy geosciences (4), pp. 103-116
4. Bentley, L.R., Zhang, J.J., and Lu, H. (1999). Four-D seismic monitoring feasibility: the CREWES report, p.777-786
5. Bodine, J.H. (1984). Waveform analysis with seismic attributes: The 54th SEG annual meeting, December, Atlanta, Expanded Abstract, p.505-509.
6. Bracewell, R.N, (1978). Fourier transform and its applications, New York: McGraw-Hill book co., Inc.
7. Brown, A.R (1996). Seismic attribute and the classification: The Leading Edge; vol. 15, p.1090.
8. Chao, W., Mian, C. and Yan, J. (2009). A prediction model for borehole stability based on seismic attribute technology: Journal of petroleum science and engineering. Vol.65, p. 208-216.

9. Cramer, Harold and Leadbeter, M.R (1967). Frequency detection and related topics: stationary and related stochastic processes, ch. 14, New York: J. Wiley and sons.
10. Gabor, D. (1946). Theory of communication, part 1, J. Inst. of electrical Eng., vol.93 part 3, p.429-441.
11. Kearey, P and Brooks, M. (1991). Introduction to exploration geophysics 2nd edition, Blackwell science publications, Oxford
12. Meckel, L.D and Nath, A.K. (1977). Geological consideration for stratigraphic modeling and interpretations. In: Payton, C.E(ed), seismic stratigraphy- Application to hydrocarbon exploration. AAPG memoir 26, 417-38
13. Obiekezie, T. N and Omoja, U. C (2019). Application of 3D Seismic Attribute Analysis for hydrocarbon prospectivity in 'Uzot-Field', On-Shore Niger Delta Basin, Nigeria. International Journal of Geophysics, volume 2019. https://doi.org/10.1155/2019/1706416
14. Ologe, O. and Olowokere, M. T. (2021). Seismic Attribute analysis a Precursor to Hydrocarbon Indicators: A Case Study of 'OK-Field', Niger Delta. Tanzania Journal of Science, 47(1): 134-144
15. Opara, A. I and Osaki, L. J (2018). 3-D Seismic Attribute for Enhanced Prospect Definition of 'Opu-Field', Coastal Swamp Depobelt Niger Delta, Nigeria, Journal of Applied Sciences vol. (18): 86-102. Doi:10.3923/jas.2018.86.102
16. Oppenheim, A.V., and Schafer, R.W. (1975). Digital signal processing: Englewood cliffs, N.J.: prentice Hall.
17. Robertson, J.D. and David A. Fisher (1988). Complex seismic trace attributes, The Leading Edge, v.7, no.6, p.22-26.Sheriff, R.E (1980). Seismic Stratigraphy. IHRDC, Boston.
18. Sheriff, R.E. (1975), Factors affecting seismic amplitude: Geophysical prospecting, v.23, p.125 – 136.
19. Sheriff, R.E. (1978). A first course in geophysical exploration and interpretation: International Human Resource Development Corporation, Boston.
20. Sheriff, Robert E. and Geldart, Lloyed P.,(1995). Exploration seismology, Cambridge: Cambridge university press.
21. Tarner, M.T., Koehler,F. And Sheriff, R.E.,(1979). Complex seismic trace analysis, Geophysics, vol.44, n.6, p.1041-1063.
22. Tokunbo, S. F., Michael, A. A., Olufemi, E. O., Opeyemi, J. A., Joel, O. A., Oluseun, A. S., Olaide, S. H., and Ajibola, R . O (2021). Hydrocarbon Reservoir Characterization of "Otan-Ile" field, Niger Delta; journal of petroleum exploration and production technology, Springer publications
23. Yao, T, and Chopra, A. (200). Integration of seismic attribute maps into 3-D facies modeling: Journal of petroleum sc. and Eng.; vol.27, p.69-84.
24. Zhang, J.J and Bentley, R.L. (2008). Complex seismic trace analysis and its application to time-lapse seismic surveys, CREWES research reports: vol.12:200.

## APPENDIX

### *Interpretation Workflow From Results*

1. Start = input/load data
2. Display conventional seismic section
3. Display polarity seismic section
4. Pick event with polarity range (4,575–11,263)
5. Define event top and bottom using velocity information
6. Evaluate thickness of event by subtraction
7. Blow event to a workable size to enhance details
8. Compute and display true seismic amplitudes of events
9. Digitize true amplitudes of event

- If true amplitude is greater than 19,764 then infer gas
- If true amplitude is between 13,840 – 17,789 then infer oil
- If amplitude is less than 13,840 then output result as unfavorable

10. Compute and display frequency section of event
11. Digitize frequency of event
    - If digital frequency is greater than 2.0 then output result as unfavorable
    - If frequency is between 18.6 – 15.0Hz then infer oil
    - If frequency is less than 15.0Hz then infer gas
12. Integrate results and make final interpretation/ recommendation
13. Tie result to spatial coordinates, post to seismic grids and save
14. End

*This page is intentionally left blank*

# Outlier Detection Procedures in a Sample from a Gumbel Distribution with Unknown Scale Parameter

*Pratyasha Tripathi & S. Lalitha*

*Tilka Manjhi Bhagalpur University*

## ABSTRACT

Lalitha and Tripathi (2018) have suggested a test statistic for the detection of a pair of outliers in a sample from a Gumbel distribution with known scale parameter σ. The statistic is not found to be suitable while dealing with the case of unknown scale parameter σ. Thus, in this paper, the test statistic suggested by Lalitha and Tripathi (2018), is suitably modified by using the modified moment estimator of the scale parameter σ for the detection of two single (upper/lower) outlying observations and one more test statistic is suggested for the detection of a pair of outlying observations. Their critical values and performance probabilities are obtained at different levels of significance by a simulation technique.

*Keywords:* extreme value theory, gumbel distribution, contaminant observation, simulation, moment estimator, slippage alternative, trimmed sample.

*Classification:* LCC Code: QA273.6

*Language:* English

Great Britain Journals Press

LJP Copyright ID: 925683
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 23 | Issue 19 | Compilation 1.0

# Outlier Detection Procedures in a Sample from a Gumbel Distribution with Unknown Scale Parameter

Pratyasha Tripathi[α] & S. Lalitha[σ]

## ABSTRACT

*Lalitha and Tripathi (2018) have suggested a test statistic for the detection of a pair of outliers in a sample from a Gumbel distribution with known scale parameter σ. The statistic is not found to be suitable while dealing with the case of unknown scale parameter σ. Thus, in this paper, the test statistic suggested by Lalitha and Tripathi (2018), is suitably modified by using the modified moment estimator of the scale parameter σ for the detection of two single (upper/lower) outlying observations and one more test statistic is suggested for the detection of a pair of outlying observations. Their critical values and performance probabilities are obtained at different levels of significance by a simulation technique.*

*Keywords:* extreme value theory, gumbel distribution, contaminant observation, simulation, moment estimator, slippage alternative, trimmed sample.

*Author* α: University Department of Statistics and Computer Applications, Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar, India.
σ : Retd. Professor, Department of Statistics, University of Allahabad, Prayagraj, Uttar Pradesh, India.

## I. INTRODUCTION

Lalitha and Tripathi (2018) have discussed detection of a single upper and lower outlier when the location parameter $\mu$ and the scale parameter $\sigma$ of Gumbel distribution were assumed to be known. But if the scale parameter $\sigma$ is not known then in such situation, previously discussed procedures should be suitably modified. Thus, when the scale parameter $\sigma$ is not known, an estimator of the scale parameter is used in the test statistic. Then it's critical values and the corresponding performance study were done by simulation technique. Since the study is about outlying observation, the entire sample should not be considered for the estimation of the scale parameter. Hence, the best linear unbiased estimate suggested by Balakrishnan and Cohen (1991) for a type II censored sample was used as an estimator for the scale parameter. But with this estimator, the values of the tabulated coefficients were available only for samples of size at most 10. Hence, a modified form of the moment estimator is considered and the statistics were studied. But on using an estimator for the scale parameter, derivation of the theoretical probability distribution of the test statistic is extremely tedious. Hence, the critical values as well as the performance of the test statistic were obtained by using simulation technique.

## II. TEST STATISTIC USING THE BEST LINEAR UNBIASED ESTIMATOR FOR THE SCALE PARAMETER

Three test statistic $Z_1'$ , $Z_2'$ and $Z_3'$ have been suggested to detect an upper outlier, lower outlier and a pair of outlying observations respectively to test the null hypothesis $H_0$ against the slippage alternative $\overline{H}$ *i.e.* there is one or a pair of observation(s) from another Gumbel with a shifted scale parameter $\sigma$ , where the scale parameter $\sigma$ was unknown. Hence for $\sigma$, the best linear unbiased estimator suggested

by Balakrishnan and Cohen (1991) given as $\sigma^* = \sum_{i=r+1}^{n-s} b_i x_{(i)}$, where $r$ and $s$ denotes number of trimmed observations from lower and upper side respectively and $b_i$ 's are known constants obtained by Balakrishnan and Chan (1992), was used. Since values of $b_i$ 's obtained by Balakrishnan and Chan (1992) up to sample size 30 but only up to 10 values are available, therefore the use of the test statistic, suggested in this case is restricted up to sample size 10. The test statistics for an upper, lower and a pair of observations obtained respectively for this case are as follows.

$$Z_1' = \frac{x_{(n)} - x_{(n-1)}}{\sigma^*}, \; Z_2' = \frac{x_{(2)} - x_{(1)}}{\sigma^*}, \; Z_3' = \frac{x_{(n)} - x_{(1)}}{\sigma^*},$$

where $x_{(1)}$ , $x_{(n-1)}$ , and $x_{(n)}$ are first, $(n\text{-}1)^{\text{th}}$ and $n^{\text{th}}$ order statistics respectively arranged in an ascending order of magnitude. However, because of the limitations in obtaining the values of $b_i$'s, these statistics have only limited usage and hence a modified form of these statistics is suggested in this paper.

### 2.1 Modified Test Statistics using a moment estimator for the scale parameter

The moment estimator of $\sigma$ was suggested by Johnson et.al. (1994) and is given as $\frac{\sqrt{6}}{\pi} s$, where $s$ is sample standard deviation, can be used as an efficient estimator of $\sigma$. This is because Johnson *et.al.* (1994) have shown that the moment estimator is about 55% more efficient than Cramer-Rao lower bound estimator of scale parameter. Further, as our work is concerned with the extreme events and the sample standard deviation is affected by the extreme observations, therefore to make the test statistic more efficient, the modified sample standard deviation $s^*$ obtained from a trimmed sample ( *i.e.* a sample obtained after deleting the extreme observations $x_{(1)}$ and $x_{(n)}$) was used instead of a complete sample standard deviation. Thus, in this case, scale parameter $\sigma$ was replaced by its modified moment estimator, given as $\sqrt{6}\, \pi\, s^*$ , where $s^*$ obtained from a trimmed sample (*i.e.* a sample obtained after deleting the extreme observations $x_{(1)}$ and $x_{(n)}$). The test statistics so obtained are as follows

$Z_1 = \frac{x_{(n)} - x_{(n-1)}}{\frac{\sqrt{6}}{\pi} s^*}$, $Z_2 = \frac{x_{(2)} - x_{(1)}}{\frac{\sqrt{6}}{\pi} s^*}$ , $Z_3 = \frac{x_{(n)} - x_{(1)}}{\frac{\sqrt{6}}{\pi} s^*}$ , where $x_{(1)}$ , $x_{(n-1)}$, and $x_{(n)}$ are first, $(n\text{-}1)$th and $n$th order statistics respectively arranged in an ascending order of magnitude. These three test statistics can be applied to a sample of size $n$= 3, 4, ..., 50. Performances of these three statistics were studied by simulation technique using 10,000 replications.

## III. CRITICAL VALUES FOR THE TEST STATISTIC Z1 TO DETECT AN UPPER OUTLIER

The test statistic $Z_1$ was used to detect an upper outlying observation in a sample from Gumbel distribution. In the detection of an upper outlying observation, the null hypothesis $H_0$ would state that there is no outlying observation in the sample. As the statistic $Z_1$ is based on the difference of the largest and second largest observations, this test statistic should reject the null hypothesis for large values of $Z_1$ . Thus an $\alpha$ - level critical region will be given as $Z_1 > z_\alpha$ where $z_\alpha$ can be obtained from $(Z_1 \geq z_\alpha) = \alpha$, where $0 < \alpha < 1$. Critical values of the test statistic $Z_1$ for case-I and case-II of the Gumbel distribution were obtained using simulation technique with 10,000 replications which are tabulated in Table 3.1. and Table.3.2. respectively for $n$ = 3(1)10, 15, 20, 30, 50 at 1%, 5% and 10% significance levels. Just as the critical values obtained for a single upper, single lower and a pair of outliers with known scale

parameter are free from the scale parameter, calculation of these critical values are also free from the scale parameters.

*Table 3.1:* Critical values of $Z_1$ for case-I at different levels of significance.

| $Z_\alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| 3 | 4.571643 | 3.087316 | 2.347163 | 11 | 4.424306 | 3.054917 | 2.289501 |
| 4 | 4.997884 | 3.171270 | 2.567949 | 12 | 4.410329 | 3.022891 | 2.254245 |
| 5 | 4.828451 | 3.170263 | 2.382530 | 13 | 4.408735 | 3.064020 | 2.237309 |
| 6 | 4.799659 | 3.146420 | 2.358684 | 14 | 4.401637 | 3.120471 | 2.224776 |
| 7 | 4.744253 | 3.137973 | 2.341062 | 15 | 4.358655 | 3.109670 | 2.209747 |
| 8 | 4.730947 | 3.120350 | 2.315699 | 20 | 4.374467 | 2.987843 | 2.126154 |
| 9 | 4.695862 | 3.111272 | 2.313943 | 30 | 4.767622 | 3.248294 | 2.304326 |
| 10 | 4.562887 | 3.099325 | 2.304717 | 50 | 4.860072 | 3.294921 | 2.579111 |

*Table 3.2.* Critical values of $Z_1$ for case-II at different levels of significance

| $Z_\alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| 3 | 2.916363 | 2.122524 | 1.794401 | 11 | 1.3700130 | 1.0359246 | 0.8400556 |
| 4 | 2.916363 | 2.122524 | 1.794401 | 12 | 1.3865416 | 1.0421608 | 0.8221364 |
| 5 | 2.532788 | 1.806103 | 1.495726 | 13 | 1.2951216 | 0.9714110 | 0.7671696 |
| 6 | 2.181685 | 1.596365 | 1.277108 | 14 | 1.1904443 | 0.9582532 | 0.7770338 |
| 7 | 1.783407 | 1.349930 | 1.136774 | 15 | 1.2876176 | 0.8898045 | 0.7190456 |
| 8 | 1.637224 | 1.193769 | 1.004956 | 20 | 1.1319771 | 0.8037215 | 0.6639349 |
| 9 | 1.4375288 | 1.0754833 | 0.9080605 | 30 | 1.0353773 | 0.7409027 | 0.6022329 |
| 10 | 1.4479635 | 1.0517048 | 0.8511008 | 50 | 0.8822868 | 0.6406025 | 0.5248657 |

## IV. CRITICAL VALUES FOR THE TEST STATISTIC TO DETECT A LOWER OUTLIER

It was observed in Lalitha and Tripathi (2018) that, for Gumbel distribution, the density functions on the real line of case-I and case-II of lower outlier are the mirror images of the density functions of the case-II and case-I of upper outlier respectively. Thus for the test statistic $Z_2$, critical values obtained in

section 4 for the upper outlier of case-I and case-II, can be used for detection of the lower outlier of case-II and case-I respectively.

## V. THE CRITICAL VALUES FOR THE TEST STATISTIC TO DETECT A PAIR OF OUTLIERS

The critical values of the test statistic $Z_3$ were obtained by using the simulation technique with 10,000 replications and are given in Table 5.1. The critical values of the test statistic $Z_3$ (unknown scale parameter) for case-I and case-II of the Gumbel are close to the critical values obtained by Lalitha and Tripathi (2018) for the Gumbel distribution case-I and case-II (known scale parameter).

*Table 5.1:* Critical values of $Z_3$ for case-I at different levels of significance.

| $Z_\alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | α = 0.01 | α = 0.05 | α = 0.10 | $n$ | α = 0.01 | α = 0.05 | α = 0.10 |
| 3 | 6.417628 | 4.963721 | 4.173418 | 11 | 7.980535 | 6.453346 | 5.588609 |
| 4 | 6.728316 | 5.020910 | 4.240807 | 12 | 8.072191 | 6.481930 | 5.838556 |
| 5 | 6.961524 | 5.206854 | 4.584614 | 13 | 8.101428 | 6.751044 | 6.060688 |
| 6 | 7.220686 | 5.614050 | 4.877048 | 14 | 8.260055 | 6.763856 | 6.063707 |
| 7 | 7.360321 | 5.802835 | 5.092647 | 15 | 8.309106 | 6.827493 | 6.055398 |
| 8 | 7.487779 | 6.061740 | 5.298490 | 20 | 8.677032 | 7.264750 | 6.477429 |
| 9 | 7.581368 | 6.097397 | 5.398235 | 30 | 8.943415 | 7.591082 | 7.051505 |
| 10 | 7.868612 | 6.195665 | 5.584236 | 50 | 9.998334 | 8.287966 | 7.574812 |

## VI. WHEN THE OBSERVATION OF A SAMPLE IS ON EITHER SIDE OF THE LOCATION PARAMETER

In this case it is assumed that some of the observations are below and some are above the location parameter. In this section the moment estimator of $\sigma$, suggested by Johnson et. al. (1994) is used in place of the scale parameter $\sigma$. The detection of an upper and a lower outlying observation can be done as discussed in section 4and section 5 respectively. But while dealing with a pair of outlying observations the procedure would be slightly different and it is given as follows.

Let $X_1, \cdots, X_n$ be a random sample from a Gumbel distribution with location parameter $\mu$ and scale parameter $\sigma$ (unknown), in which some, say $m$, observations, $X_1, \cdots, X_m$ are less than the location parameter $\mu$ and rest of the $(n - m)$ observations $X_{m+1}, \cdots, X_n$ of the sample are greater than $\mu$. Then considering the $m$ observations lying below the location parameter, a modified test statistic can be defined for detecting the smallest observation by considering $m^{\text{th}}$ largest observation as $X_{(n)}$, *i.e.* $Z =$

$$\frac{X_{(m)} - X_{(1)}}{\frac{\sqrt{6}}{\pi} S^*}, m = 2, \cdots, n-1.$$

As before, the observation corresponding to $X_{(m)}$ cannot be declared as an outlying one, being the observation lying closest to the location parameter, when the test statistic falls in the critical region. Thus, in the event of rejection of the null hypothesis, only the smallest observation *i.e.* $X_{(1)}$ should be

considered as outlying observation. In this case, the critical values given in case II of section 4 should be used for testing the null hypothesis with a sample size as $m$. For the rest, $(n - m)$ observations above the location parameter, a modified test statistic can be defined for detecting the largest observation by considering $(m + 1)^{th}$ largest observation as $X_{(1)}$, *i.e.*

$$Z = \frac{X_{(n)} - X_{(m+1)}}{\frac{\sqrt{6}}{\pi} s^*}, m = 1, \cdots, n - 2.$$

Here again, the observation corresponding to $X_{(m+1)}$ would be lying closest to the location parameter and therefore this observation cannot be declared as an outlying one. In this case the critical values obtained in case I of section 4 should be used for testing the null hypothesis with a suitable modification of the sample size as $(n - m)$. As before, in the event of rejection of the null hypothesis, the largest observation *i.e.* $X_{(n)}$ should be declared as outlying observation.

*a. Case when only one observation is on either side of the location parameter and the location parameter is also known*

(*i*). When only one observation is lying below the location parameter which is assumed to be known, while all other observations are above the location parameter, then the statistic $Z$ can be modified as

$$Z = \frac{\mu - X_{(1)}}{\frac{\sqrt{6}}{\pi} s^*}. \quad (6.1)$$

Here, if it is assumed that $X_{(1)} < \mu$ and $X_{(1)}$ follows a Gumbel distribution as defined in case II with location and scale parameters $\mu$ and $\sigma$ respectively. The critical values $z_\alpha$ can be obtained by using simulation technique with 10,000 replicates. These critical values were tabulated in Table 6.1 for different values of sample size $n$ and different levels of significance, given as follows.

*Table 6.1:* The critical values $z_\alpha$ of the test statistic when only one observation is lying below the location parameter (known)

| $z_\alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| 5 | 1.80287 | 0.92473 | 0.65052 | 12 | 0.65225 | 0.45047 | 0.40411 |
| 6 | 1.16972 | 0.64228 | 0.48095 | 13 | 0.64222 | 0.44603 | 0.41108 |
| 7 | 1.09230 | 0.58027 | 0.44450 | 14 | 0.68654 | 0.45007 | 0.41784 |
| 8 | 0.88634 | 0.55103 | 0.42177 | 15 | 0.73676 | 0.47775 | 0.42694 |
| 9 | 0.73802 | 0.49250 | 0.39991 | 20 | 0.63360 | 0.48989 | 0.46042 |
| 10 | 0.72427 | 0.47692 | 0.93198 | 30 | 0.61697 | 0.54320 | 0.51112 |
| 11 | 0.68366 | 0.43731 | 0.38694 | 50 | 0.68311 | 0.60397 | 0.55898 |

(*ii*). When only one observation is lying above the location parameter, while all other observations are below the location parameter, then the statistic Z is modified as

$$Z = \frac{X_{(n)} - \mu}{\frac{\sqrt{6}}{\pi} s^*}. \quad (6.2)$$

Here, as it is assumed that $\mu < X_{(n)}$ , $X_{(n)}$ follows a Gumbel distribution as defined in case I with location and scale parameters $\mu$ and $\sigma$ respectively. The critical values $z_{\alpha}$ can be obtained by using simulation technique with 10,000 replications and were tabulated in Table 6.2., as given below

*Table 6.2:* The critical values $z_{\alpha}$ of the test statistic when only one observation is lying above the location parameter (known)

| $z_{\alpha}$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | α = 0.01 | α = 0.05 | α = 0.10 | $n$ | α = 0.01 | α = 0.05 | α = 0.10 |
| 5 | 1.81857 | 0.92633 | 0.60187 | 12 | 0.65055 | 0.46036 | 0.40235 |
| 6 | 1.25745 | 0.6888$^{0}$ | 0.49004 | 13 | 0.67775 | 0.44593 | 0.40829 |
| 7 | 1.04826 | 0.58507 | 0.45963 | 14 | 0.66629 | 0.45812 | 0.42150 |
| 8 | 0.94477 | 0.56933 | 0.41839 | 15 | 0.67564 | 0.46658 | 0.42315 |
| 9 | 0.91680 | 0.48056 | 0.38863 | 20 | 0.55692 | 0.48612 | 0.45476 |
| 10 | 0.75254 | 0.49148 | 0.40449 | 30 | 0.60539 | 0.54456 | 0.50122 |
| 11 | 0.74374 | 0.49846 | 0.40382 | 50 | 0.69922 | 0.60899 | 0.56216 |

*b. Case when the location parameter is unknown*

When the location parameter is unknown, it can be estimated with a trimmed sample, as suggested in section 3 and is denoted by $\mu^{*}$. As the two extreme observations are suspected outliers therefore the sample should be trimmed at both the ends. The location parameter used in the test statistics given in equations (6.1) and (6.2), is replaced by this estimate $\mu^{*}$ . With this estimate, the number of observations on its left and right, *i.e.* $m$ and $n$-$m$ can be decided. Then for $m$ >1and/ $(n - m)$ >1, the above said procedures can be used, as their critical values are independent of both the location and scale parameter. Also when $m$ =1 and/$(n - m)$ =1, the test statistic will be as given above in section 6(a), with the value of the location parameter replaced by its estimator $\mu^{*}$ . The critical values so obtained were tabulate in Table 6.3. and Table 6.4. for the two cases of the distribution *i.e.* when only one observation is below the location parameter and when only one observation is above the location parameter respectively, are as given below.

*Table 6.3:* The critical values $z\alpha$ of the test statistic when only one observation is lying below the location parameter (unknown)

| $z_{\alpha}$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | α = 0.01 | α = 0.05 | α = 0.10 | $n$ | α = 0.01 | α = 0.05 | α = 0.10 |
| 5 | 1.66908 | 0.68764 | 0.47271 | 12 | 0.61636 | 0.38536 | 0.33359 |
| 6 | 1.35244 | 0.65402 | 0.44787 | 13 | 0.57881 | 0.40841 | 0.35085 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 7 | 1.00532 | 0.54035 | 0.38811 | 14 | 0.65151 | 0.41213 | 0.36337 |
| 8 | 1.02801 | 0.47312 | 0.36072 | 15 | 0.67439 | 0.39163 | 0.36625 |
| 9 | 0.75887 | 0.45592 | 0.33152 | 20 | 0.57011 | 0.42248 | 0.39505 |
| 10 | 0.87499 | 0.46345 | 0.33778 | 30 | 0.56384 | 0.47950 | 0.44343 |
| 11 | 0.70889 | 0.45018 | 0.33154 | 50 | 0.64258 | 0.54803 | 0.50708 |

*Table 6.4:* The critical values $z\alpha$ of the test statistic when only one observation is lying above the location parameter (unknown)

| $z_\alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ | $n$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
| 5 | 2.37667 | 0.96049 | 0.62846 | 12 | 0.70681 | 0.40361 | 0.33457 |
| 6 | 1.43159 | 0.63315 | 0.63315 | 13 | 0.58528 | 0.41091 | 0.34473 |
| 7 | 0.94607 | 0.42759 | 0.33917 | 14 | 0.67370 | 0.39965 | 0.35868 |
| 8 | 0.77505 | 0.42759 | 0.33917 | 15 | 0.62877 | 0.39552 | 0.36069 |
| 9 | 0.79099 | 0.44493 | 0.33463 | 20 | 0.56641 | 0.44057 | 0.40532 |
| 10 | 0.69059 | 0.42149 | 0.31121 | 30 | 0.55085 | 0.48315 | 0.44851 |
| 11 | 0.70327 | 0.39549 | 0.32351 | 50 | 0.62877 | 0.54954 | 0.50829 |

Here the critical values were obtained up to sample size 50, therefore the suggested test statistic, based on the trimmed estimate of the scale parameter can be used for large samples as well.

## VII. PERFORMANCE STUDY FOR A SINGLE UPPER OUTLIER

Consider a set of observations which contains a single contaminant observation $x_c$ *and which comes from a population that* has a different distribution from the rest of the observations. The power of the test can be defined as $P[Z_n > z_\alpha|$the sample contains one contaminant observation], *i.e.* the probability of the largest observation of the sample is being identified as discordant. But it is not necessary that the test always identifies $x_c$ as a discordant or defining only power of a test statistic is not sufficient to describe performance of a test statistic. Barnett and Lewis (1994) have suggested probabilities, $P_1$ to $P_3$ as test performance criteria. To improve the performance criteria of a statistic, David and Nagaraja (2003) discussed the properties of five such probabilities, labeled as $P_1$ to $P_5$ as a reasonable measure of the performance of $Z_n$. These probabilities are defined as

$P_1 = \Pr[Z_n > z_\alpha|\overline{H}]$, is the probability the observation tested by the test statistic is identified as the outlying one, when it is known that there is one outlier is present.

$P_2 = \Pr[Z_c > z_\alpha|\overline{H}]$, is the probability the contaminant observation tested by the test statistic is identified as the outlying one, when it is known that there is one outlier is present.

$P_3$ = Pr $[Z_n = Z_c,\ Z_n > z_\alpha|\overline{H}]$, is the probability the contaminant observation is the extreme observation tested by the test statistic which is identified as the outlying one, when it is known that there is one outlier is present.

$P_4$ = Pr $[Z_n = Z_c > z_\alpha,\ Z_{(n-1)} < z_\alpha|\overline{H}]$ is the probability that the largest observation is the contaminant observation, which is being detected as discordant while the second largest is not a discordant observation; and

$P_5$ = Pr $[Z_c > z_\alpha|Z_n = Z_c; \overline{H}]$, is the conditional probability that when it is given that the contaminant is an outlier and it is also identified as discordant by the test, where $Z_n$ is a general test statistic and $Z_c$ is the corresponding value for the contaminant observation. David and Nagaraja (2003) have observed that $P_{1\geq}P_{2\geq}P_{3\geq}P_4$ and the information given by $P_2$ and $P_4$ are seen to be limited. Thus, only $P_1$, $P_3$ and $P_5$ are sufficient for defining performance of any test statistic. Further, Hayes and Kinsella (2003) discussed the performance of the discordancy test on the basis of six performance criteria and called them as non spurious power, spurious power, swamping effect, spurious Type II error, partially spurious Type II error and nonspurious Type II error. Also, it was suggested that a good discordancy test should have high value of non spurious power, low value of spurious power and low value of swamping effect. They recommended that out of six performance criteria, probability of the non-spurious power $P_3$, probability of the spurious power$P_1$ and the probability of the non-spurious type-II error $P_1$-$P_3$ (which gives the probability that the test wrongly identifies a good observation as discordant), are required to specify the test completely. In accordance with the above given probabilities, large values of $P_5$ and $P_3/P_5$- the probability that the contaminant shows up as the outlier, are desired.

In this section, study of the performance for detection of a lower outlier for the both the cases of Gumbel distribution with location parameter $\mu$ and scale parameter $\sigma$ (unknown) are discussed. Case-I: When all observations are greater than the location parameter $\mu$. The performance of the test statistic $Z_1$ to detect an upper outlier in a sample from a Gumbel distribution, after introducing a contaminant observation from another sample from the same distribution with different scale parameter, was observed. This was replicated 10,000 times. From this, the values of the probabilities $P_1$, $P_3$, $P_5$, $P_1$- $P_3$ and $P_3/P_5$ for the test statistic $Z_1$ were obtained and are given in Table 7.1. It can be observed that at 1% level of significance, the power $P_1$, the non- spurious power $P_3$ and the conditional power $P_5$ of the test statistic are showing almost negligible changes between sample size 5 and 10, between sample sizes 10 and 15 a rapid increase is observed while beyond 15 the rate of increase is comparatively low. The non-spurious type-II error is also showing almost no change between sample size 5 and 10, between sample sizes 10 and 15 it is increasing while beyond 15 it decreases. The ratio $P_3/P_5$ *i.e.* the probability that the contaminant shows up as an outlier, is almost constant between sample sizes 5 and 10, between sample sizes 10 and 15 it increases with very high rate of increase and beyond 15 rate of increase is comparatively low. Thus, it can be interpreted that the performance of the test statistic is increasing with sample size but beyond sample size 20 there is hardly any variation.

*Table 7.1:* Performance of the test statistic $Z_1$ for case-I at different levels of significance

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.51490 | 0.42407 | 0.09083 | 0.55610 | 0.76258 |
| | 5% | 0.65360 | 0.53636 | 0.11724 | 0.70050 | 0.76568 |
| | 10% | 0.73890 | 0.70982 | 0.02908 | 0.86990 | 0.81598 |
| | 1% | 0.51490 | 0.42407 | 0.09083 | 0.53510 | 0.79251 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | 5% | 0.90880 | 0.84694 | 0.06186 | 0.90050 | 0.94052 |
| | 10% | 0.94860 | 0.91699 | 0.03161 | 0.94040 | 0.97511 |
| 15 | 1% | 0.94009 | 0.75653 | 0.18356 | 0.83550 | 0.90548 |
| | 5% | 0.97690 | 0.92238 | 0.05452 | 0.94980 | 0.97113 |
| | 10% | 0.99020 | 0.95652 | 0.03368 | 0.96950 | 0.98661 |
| 20 | 1% | 0.95090 | 0.89512 | 0.05578 | 0.93330 | 0.95909 |
| | 5% | 0.99200 | 0.95667 | 0.03533 | 0.96990 | 0.98636 |
| | 10% | 0.99820 | 0.99500 | 0.00320 | 0.99840 | 0.99659 |

It can be seen that at 5% level of significance the power, the non spurious power, the conditional power and the probability that the contaminant is identified as an outlier all are increasing rapidly between sample sizes 5 and 10 and beyond 10 its rate of increase is comparatively low. The values of all these probabilities are very high for sample sizes up to 15 while beyond 15 it shows almost no fluctuation. The non spurious type-II error is decreasing very fast between sample sizes 5 and 10 while beyond that it decreases with comparatively low rate of decrease. Hence from inferences discussed above it can be concluded that the performance of the test statistic is good at 5% level of significance. It can be noticed from. at 10% level of significance the power, the non spurious power, the conditional power and the probability of showing up a contaminant as an outlier are very high and increasing rapidly between sample sizes 5 and 10 and beyond that its rate of increase is very low. The non spurious type-II error is increasing with a very low rate of increase with the increase of the sample size up to 15 and beyond sample size 15 it decreases slowly. Hence it can be seen that the calculated probabilities $P_1$, $P_3$, $P_5$ and $P_3/P_5$ are high and $P_1$- $P_3$ *i.e.* non spurious error is low, as desired. Thus, on the basis of above results it can be interpreted that the performance is increasing with sample size up to a certain level and after that the variations are almost uniform, also the performance of the statistic is found to be good for the detection of an upper outlier in a Gumbel sample at 5% and 10% levels of significance, while at 1% level it is comparatively low.

Case-II: When all observations were smaller than the location parameter $\mu$. The performance of the test statistic $Z_1$to detect an upper outlier in a sample from a Gumbel distribution, after introducing a contaminant observation from another sample of the same distribution with different scale parameter was obtained. These performance probabilities were obtained by simulation technique with 10,000 replicates and are given in

*Table 7.2:* Performance of the test statistic $Z_1$ for case-II at different levels of significance

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.3787 | 0.3488 | 0.0299 | 0.5605 | 0.6222 |
| | 5% | 0.5284 | 0.4918 | 0.0366 | 0.7017 | 0.7009 |
| | 10% | 0.6074 | 0.5407 | 0.0667 | 0.7638 | 0.7079 |
| 10 | 1% | 0.8716 | 0.8071 | 0.0645 | 0.8961 | 0.9007 |
| | 5% | 0.8913 | 0.8358 | 0.0554 | 0.9029 | 0.9257 |
| | 10% | 0.9410 | 0.9090 | 0.0319 | 0.9263 | 0.9813 |
| 15 | 1% | 0.9840 | 0.9329 | 0.0511 | 0.9484 | 0.9837 |
| | 5% | 0.9912 | 0.9447 | 0.0464 | 0.9583 | 0.9858 |
| | 10% | 0.9944 | 0.9762 | 0.0182 | 0.9874 | 0.9887 |
| 20 | 1% | 0.9983 | 0.9500 | 0.0483 | 0.9729 | 0.9765 |
| | 5% | 0.9971 | 0.9691 | 0.0279 | 0.9888 | 0.9801 |
| | 10% | 0.9991 | 0.9866 | 0.0125 | 0.9912 | 0.9953 |

It was observed from Table 7.2. that the performance of the test statistic under consideration was found to be good (as the performance is increasing with the increase of the sample size and the calculated probabilities $P_1$, $P_3$, $P_5$ and $P_3/P_5$ are high and $P_1$- $P_3$ is low, as desired). Also, it can be seen that at 1% level of significance the power, the nonspurious power, the conditional power and the probability of identifying a contaminant as an outlier, are increasing rapidly between sample sizes 5 and 10, beyond that the changes are almost negligible. The non-spurious type-II error is increasing with a small rate of increase at 1% and 5% levels while at 10% level of significance it is decreasing with a very small rate of decrease. Thus, it can be interpreted that the performance of the statistic is increasing with the sample size up to 15 and beyond that almost negligible changes are observed.

## VII. PERFORMANCE STUDY FOR A SINGLE LOWER OUTLIER

In this section the performance of $Z_2$ was studied for both the cases of the Gumbel distribution. Case-I: When all observations were greater than the location parameter $\mu$. The performance of the test statistic $Z_2$ to detect a lower outlier in a Gumbel sample, after introducing a contaminant observation from another sample of the same distribution with different scale parameter, was studied by simulation technique with 10,000 replicates. These performance probabilities are given in Table 8.1. From this table, it can be observed that the performance is increasing with the increase of the sample size and it is found to be satisfactory beyond sample size 15.

*Table 8.1:* Performance of the test statistic $Z_2$ for case-I at different levels of significanc

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.5704 | 0.5124 | 0.0580 | 0.7077 | 0.7240 |
| | 5% | 0.6532 | 0.6331 | 0.0201 | 0.7243 | 0.8741 |
| | 10% | 0.7214 | 0.6865 | 0.0349 | 0.7456 | 0.9207 |
| 10 | 1% | 0.7360 | 0.6791 | 0.0569 | 0.9276 | 0.7321 |
| | 5% | 0.8222 | 0.7955 | 0.0267 | 0.9219 | 0.8629 |
| | 10% | 0.9381 | 0.9087 | 0.0294 | 0.9478 | 0.9587 |
| 15 | 1% | 0.8139 | 0.756 | 0.0579 | 0.991 | 0.7629 |
| | 5% | 0.9684 | 0.9359 | 0.0325 | 0.9956 | 0.9400 |
| | 10% | 0.9842 | 0.955 | 0.0292 | 0.9956 | 0.9592 |
| 20 | 1% | 0.9512 | 0.8965 | 0.0547 | 0.9991 | 0.8973 |
| | 5% | 0.9878 | 0.9503 | 0.0375 | 0.9994 | 0.9509 |
| | 10% | 0.9975 | 0.9699 | 0.0276 | 0.9996 | 0.9703 |

It can be seen that the power $P_1$, the non spurious power $P_3$ and the conditional power $P_5$ are increasing rapidly between sample sizes 5 and 10 at 1% and 10% levels of significance while are increasing rapidly between sample sizes 5 and 15 at 5% level of significance. The non spurious type-II is found to be very low and it shows very small changes throughout. The probability of showing the contaminant as an outlier, is high for high values (greater than 10) of the sample size, it is increasing rapidly between sample sizes 5 and 10 at 1% level of significance. While at 5% and 10% levels of significance it is showing negligible changes between sample sizes 5 and 10, beyond 10 it increases very slowly. On the basis of above inference, it can be interpreted that the power $P_1$, the non spurious power $P_3$, the conditional power $P_5$ and the probability that the contaminant is showing up as an outlier, are found to

be very high and the non spurious type-II error is low, as desired. Thus, it can be said that the test statistic is performing very well in this case for sample sizes greater than 10.

*Case-II:* When all observations were smaller than the location parameter $\mu$. The performance of the test statistic $Z_2$ for detection of a lower outlying observation from a Gumbel sample with a contaminant observation which was taken from a sample of the same distribution with different scale parameter was studied. All the performance probabilities were calculated by simulation technique with 10,000 replications, and are given in Table 8.2. From Table 8.2. it can be seen that the performance is increasing with the sample size and it is good enough for sample sizes greater than 10.

*Table 8.2:* Performance of the test statistic $Z_2$ for case-II at different levels of significance

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.5581 | 0.4566 | 0.1015 | 0.5616 | 0.8130 |
| | 5% | 0.7862 | 0.7498 | 0.0364 | 0.7934 | 0.9450 |
| | 10% | 0.8786 | 0.8543 | 0.0243 | 0.8848 | 0.9655 |
| 10 | 1% | 0.7949 | 0.7786 | 0.0163 | 0.7987 | 0.9548 |
| | 5% | 0.9794 | 0.9364 | 0.0430 | 0.9788 | 0.9767 |
| | 10% | 0.9831 | 0.9646 | 0.0185 | 0.9837 | 0.9806 |
| 15 | 1% | 0.9370 | 0.9263 | 0.0107 | 0.9362 | 0.9694 |
| | 5% | 0.9976 | 0.9768 | 0.0208 | 0.9979 | 0.9789 |
| | 10% | 0.9987 | 0.9854 | 0.0133 | 0.9987 | 0.9867 |
| 20 | 1% | 0.9804 | 0.9760 | 0.0044 | 0.9831 | 0.9828 |
| | 5% | 0.9999 | 0.9883 | 0.0116 | 0.9979 | 0.9904 |
| | 10% | 1.0000 | 0.9981 | 0.0019 | 1.0000 | 0.9981 |

From this, it can be interpreted that the power, the non spurious power and the conditional power of the test statistic are high for large sample sizes; these are increasing rapidly between sample sizes 5 and 10 while beyond sample size 10 it increases with a comparatively low rate of increase. The non spurious type-II error is low and decreasing slowly with a very small rate of decrease. The probability of identifying the contaminant as an outlying observation is also very high and increasing with the increase of the sample size, as desired for a good discordancy test according to Barnett and Lewis (1994) and Hayes and Kinsella (2003). Thus it can be concluded that the test statistic under consideration is performing very well for sample sizes greater than 5

## IX. PERFORMANCE STUDY FOR A PAIR OF OUTLIERS

Case-I: When all observations were greater than the location parameter $\mu$. The performance of the test statistic $Z_3$ to detect a pair of outliers in a sample from Gumbel distribution after introducing a pair of contaminants from another sample from the same distribution with a shifted scale parameter was obtained by simulation technique with 10,000 replicates. These performance probabilities are given in Table 9.1.

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.2678 | 0.1790 | 0.0888 | 0.6731 | 0.2659 |
| | 5% | 0.5066 | 0.5008 | 0.0058 | 0.8499 | 0.5892 |
| | 10% | 0.6139 | 0.5360 | 0.0779 | 0.8996 | 0.5958 |
| | 1% | 0.5514 | 0.4550 | 0.0964 | 0.8736 | 0.5208 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | 5% | 0.8019 | 0.7095 | 0.0924 | 0.9665 | 0.7341 |
| | 10% | 0.8793 | 0.7779 | 0.1014 | 0.9849 | 0.7898 |
| 15 | 1% | 0.7510 | 0.6325 | 0.1185 | 0.9468 | 0.6680 |
| | 5% | 0.9169 | 0.8216 | 0.0953 | 0.9900 | 0.8299 |
| | 10% | 0.9668 | 0.8376 | 0.1292 | 0.9967 | 0.8404 |
| 20 | 1% | 0.8579 | 0.7253 | 0.1326 | 0.9774 | 0.7421 |
| | 5% | 0.9646 | 0.8512 | 0.1134 | 0.9966 | 0.8541 |
| | 10% | 0.9896 | 0.8406 | 0.1490 | 0.9990 | 0.8414 |

It can be observed from the above table that the power, the non-spurious power and the conditional power of the test statistic are increasing rapidly throughout. It can also be seen that the power, the non spurious power, the conditional power and the probability of identifying the contaminant as an outlier is increasing with the increase of the sample size and the non spurious type-II error is very low. Since the probabilities $P_1$, $P_3$, $P_5$ are significantly high for sample sizes greater than 10 and $P_1 - P_3$ is low, as desired for a good discordancy test. Thus, it can be interpreted that the performance the test statistic $Z_3$ is good for large samples.

Case-II: When all observations were smaller than the location parameter $\mu$. The performance of the test statistic $Z_3$ to detect a pair of outliers in a sample from a Gumbel distribution after introducing a pair of contaminants from another sample from the same distribution with different scale parameter was studied by simulation technique with 10,000 replications. These performance probabilities are given in Table 9.2.

*Table 9.2:* Performance of the test statistic $Z_3$ for case-II at different levels of significance.

| $n$ | Level of Significance | $P_1$ | $P_3$ | $P_1 - P_3$ | $P_5$ | $P_3/P_5$ |
|---|---|---|---|---|---|---|
| 5 | 1% | 0.0154 | 0.0066 | 0.0088 | 0.1915 | 0.0345 |
| | 5% | 0.0891 | 0.0833 | 0.0058 | 0.4489 | 0.1856 |
| | 10% | 0.2008 | 0.1229 | 0.0779 | 0.6354 | 0.1934 |
| 10 | 1% | 0.0644 | 0.0480 | 0.0164 | 0.3916 | 0.1226 |
| | 5% | 0.2741 | 0.1837 | 0.0904 | 0.7131 | 0.2576 |
| | 10% | 0.4563 | 0.4161 | 0.0402 | 0.8393 | 0.4958 |
| 15 | 1% | 0.1149 | 0.0964 | 0.0185 | 0.4867 | 0.1981 |
| | 5% | 0.3706 | 0.2753 | 0.0953 | 0.7651 | 0.3598 |
| | 10% | 0.6182 | 0.5250 | 0.0932 | 0.9028 | 0.5815 |
| 20 | 1% | 0.1526 | 0.1100 | 0.0426 | 0.4926 | 0.2233 |
| | 5% | 0.5848 | 0.4714 | 0.1134 | 0.8628 | 0.5463 |
| | 10% | 0.8107 | 0.6356 | 0.1751 | 0.9648 | 0.6587 |

From Table 9.2. it can be observed that since the power, the non spurious power, the conditional power, are very low thus the test statistic is not found to be satisfactory for the detection of a pair of outliers in a sample from case-II of a Gumbel distribution. It can be seen that the power of the statistic is increasing throughout with almost constant rate of increase. The non spurious power, is increasing with a good rate of increase between sample sizes 5 and 15 while beyond that almost no changes are observed. The non spurious type II error of the test statistic at 1% level of significance is found to be very low. The conditional power of the statistic at 1% level of significance shows that the conditional

power increases between sample sizes 5 and 10, between sample sizes 10 and 15 almost no changes are observed while beyond sample size 15 again it increases uniformly. The probability of detecting an outlying pair at 1% level of significance is found to be low. It increases rapidly between sample sizes 5 and 15, while beyond 15 the rate of increase is low.

It can also be seen that the power increases with the increase of sample size throughout almost uniformly but small changes are occurring between sample sizes 10 and 15. The non- spurious power of the test statistic at 5% level of significance increases rapidly from sample size 5 to 15 and beyond sample size15, its rate of increase is comparatively low. The spurious power of the test statistic at 5% level of significance increases rapidly between sample sizes 5 and 10, the rate of increase is comparatively low between sample sizes 10 and 25 while beyond sample size 25, almost no changes are seen. The conditional power increases with good rate of increase between sample sizes 5 and 10 but gives relatively low rate of increase between sample sizes 10 and 20, beyond sample size 20 it increases uniformly. The probability of detecting contaminant(s) as an outlying pair at 5% level of significance is found to be good and is increasing with the increase of the sample size.

It can be noticed at 10% level of significance that the power of the test statistic under consideration increases with the increase of sample size up to 15 and beyond sample size 15 it gives relatively small changes. The non- spurious power of the test statistic at 10% level of significance increases rapidly with the increase of sample size from sample size 5 to 10 and beyond sample size 10 it increases uniformly with comparatively low rate of increase. The spurious power of the test statistic at 10% level of significance is found to be low. It decreases between sample sizes 5 and 10 and then starts to increase beyond 10 with a good rate of increase. The conditional power of the statistic at 10% level of significance shows that the conditional power increases rapidly between sample sizes 5 and 20. The probability of detecting contaminant(s) as an outlying pair at 10% level of significance is found to be high and increases with the increase of the sample size up to 10, beyond 10 the rate of increase is comparatively low. It can be concluded from the above inference that the test statistic is not found to be suitable for the detection of a pair of outliers in a sample from a Gumbel distribution. Therefore, it cannot be recommended for detection of a pair of outliers in a sample from a Gumbel distribution for case-II.

*Conclusion*: It can be concluded from the above study that the two suggested test statistics $Z_1$ and $Z_2$ (for detection of the single upper and lower outlier) are performing very well, while the performance of the test statistic suggested for the detection of a pair of outliers is very poor, especially for case-II. Thus, use of the test statistics $Z_1$and $Z_2$ for detection of an upper and lower outlying observation respectively, can be used for the Gumbel distribution with unknown scale parameter but the test statistic $Z_3$ cannot be suggested for efficient results in a Gumbel distribution.

## REFERENCES


1. Balakrishnan, N. and Chan, P.S., *Extended tables of Best Linear Unbiased Estimates from complete and Type II censored samples from the extreme value distribution for sample size up to 30,* Report, Department of Mathematics and Statistics, McMaster University, Hamilton, Canada (1992).
2. Balakrishnan, N. and Cohen, A.C., *Order Statistics and Inference: Estimation Methods,* Academic Press, San Diego, CA, 1991.
3. Barnett, V. and Lewis, T., *Outliers in Statistical Data*, 3rd edition. J. Wiley & Sons 1994, XVII. 582 pp.
4. David, H. A., Nagaraja, H. N., *Order Statistics*, 3rd edition, JohnWiley & Sons, 2003

5. Hayes, K. and Kinsella, T. *Spurious and non-spurious power in performance criteria for tests of discordancy*, The Statistician 52 (2003), pp. 69–82.
6. Johnson, N. L., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distribution*, Vol. 1, 2 nd edition, New York; Wiley, 1994.
7. Lalitha, S. and Tripathi, P., *Detection of a pair of outliers in a sample from a Gumbel Distribution with known scale parameter*, Journal of Applied Statistics, (2018) 45, pp.243-254.

# On the Distribution of Gaussian Primes

*Xian Hemingway*

## ABSTRACT

In this paper, we will first explore the distribution of Gaussian primes in the first quadrant locally and, furthermore, use the findings to investigate the distribution of Gaussian primes on a wide variety of straight lines.

*Keywords:* gaussian integers; the distribution of gaussian primes; the elementary line.

*Classification:* LCC Code: QA1-QA939

*Language:* English

Great Britain Journals Press

LJP Copyright ID: 925684
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 23 | Issue 19 | Compilation 1.0

# On the Distribution of Gaussian Primes

Xian Hemingway

## ABSTRACT

In this paper, we will first explore the distribution of Gaussian primes in the first quadrant locally and, furthermore, use the findings to investigate the distribution of Gaussian primes on a wide variety of straight lines.

*Keywords and phrases:* gaussian integers; the distribution of gaussian primes; the elementary line.



## I. INTRODUCTION

We know that if a Gaussian integer $a = x + iy$ $(x, y \in \mathbb{Z})$ is a Gaussian prime without the case $a = 1 + i$ , its norm

$$N(a) = x^2 + y^2$$

is either [1] a rational prime of type $4k+1$ or the square of a rational prime of type $4k+3$ , where $k \in \mathbb{Z}^+$. The Gaussian primes corresponding to the latter must be distributed on the coordinate axes of the complex plane, while the former must be symmetrically distributed in the quadrants of the coordinate system excluding the coordinate axes.

There must be, obviously, an infinite number of Gaussian primes on the coordinate axes of the complex plane, because of the theorem derived by Dirichlet as follows.

**Lemma 1.1.** *[2]There must be infinitely many rational primes of type $an + b$, where $n \in \mathbb{Z}^+$, and $(a, b) = 1$.*

It can be seen that, with $(4, 3) = 1$, there are infinitely many rational primes of type $4k + 3$ while the required conclusion is derived [3]. Then are there other lines with infinite number of Gaussian primes? In particular, are there infinitely many Gaussian primes with its real part being 1? These problems, which are currently unsolved by mathematicians, are exactly what this paper needs to address. It also provides a basic idea for studying Gaussian primes.

## II. THE BASIC PROPOSITIONS

**Proposition 2.1.** *Let $K$ be a circle in the complex plane centered at the origin with the case that, which there exists one Gaussian integer point being the Gaussian prime on, then all Gaussian integer points on $K$ is Gaussian prime.*

*Proof.* From the geometric meaning of the norm on the complex plane, combined with the uniqueness of the rational prime $p$ decomposition into the form $x^2 + y^2$, so that it is easy to prove the proposition.

**Definition 2.2.** Let the Gaussian integer $a = x + iy$, then $a^{\top} = y + ix$ is said to be the absolute conjugate of $a$ with $xy > 0$ while $a^{\top} = -(y + ix)$ is said to be the absolute conjugate of a with $xy < 0$.

Clearly the norm of both satisfies $N(a) = N(a^{\top})$ and, if a is a Gaussian prime, then $a^{\top}$ is also a Gaussian prime. It is easy to obtain that there are no three Gaussian prime points of equal distance to the origin in the first quadrant of the complex plane.

**Proposition 2.3.** *All Gaussian primes in the first quadrant of the complex plane, except for* $1 + i$ *, are distributed on the line*

$$l_k : x + y = 2k + 1, \ k \in \mathbb{Z}^+ \tag{2.1}$$

*which is called the elementary line on the complex plane. There, namely, are no Gaussian prime points between the lines* $l_k$ *and* $l_{k+1}$ *.*

**Proposition 2.4.** *Given any positive integer* $t$ *, there exists an infinite number of the elementary lines* $l_k$ *on which there are Gaussian prime points with the real part being* $t$ *or the imaginary part being* $t$ *.*

**Proposition 2.5.** *There, from the whole complex plane, are infinitely many Gaussian primes on the line* $l : \pm x + y = 2k + 1, \ k \in \mathbb{Z}$.

**Proposition 2.6.** *Given any Gaussian integers* $a$ *and* $b$ *, there are infinitely many Gaussian primes on the line which, being different from the line* $l : \pm x + y = 2k, \ k \in \mathbb{Z}$*, is determined by the points* $a$ *and* $b$ *.*

The proofs of each of these propositions are given below. For convenience, the following Gaussian primes are referred to as primes while the primes defined over the real domain are called the rational primes and, the Gaussian integers (points) are referred to as the integer points or lattice points.

## III. PROOF OF THE PROPOSITION

### *3.1. Proof of Proposition 2:3.*

*Proof.* Apparently, all integer points in the first quadrant of the complex plane including the coordinate axes are distributed on the line $l_k : x + y = 2k - 1$ and $l_k^{'} : x + y = 2k$, where $k \in \mathbb{Z}^+$. When $k = 1$ , there are no primes on $l_k$ while there is only a unique prime $1 + i$ on $l_k^{'}$, with the norm $N(1 + i) = 2$.

Now consider the case of the distribution of primes on the line $l_k^{'} : x + y = 2k + 2, \ k \in \mathbb{Z}^+$. Easy to know that, in the event of the integer $a = x + iy$ being a prime, its norm $N(a) \equiv 1(\text{mod}4)$ ; So that the inverse negative propositions of it is that, if the norm of the integer $a = x + iy$ does not satisfy $N(a) \equiv 1(\text{mod}4)$ , $a$ is not prime. Here in evidence,

$$N(a) = x^2 + (2k + 2 - x)^2 = 2x^2 - 4(k + 1)x + 4(k + 1)^2 \equiv 0, 2(\text{mod}4) \tag{3.1}$$




London Journal of Research in Science: Natural and Formal

where it can be seen that the integer a must not be a prime and thus all primes except $1+i$ are distributed on the line $l_k : x+y = 2k+1,\ k \in \mathbb{Z}^+$ .

Further, the distribution of Gaussian primes on the elementary line $l_k$ is symmetric with respect to the line $l : y = x$.

**Corollary 3.1.** *From the symmetry of the distribution of primes, it follows that all primes in the entire complex plane are distributed on the closed quadrilateral*

$$|x| + |y| = 2k+1,\ k \in \mathbb{Z}^+ \tag{3.2}$$

*without the case of the special primes* $\pm 1 \pm i$ *on the quadrilateral* $|x| + |y| = 2$.

### 3.2. Proof of Proposition 2:4.

*Proof.* Given a positive integer $t$ , the lines where the primes $a = t+iy$ and $b = x+it$ are located is symmetric with respect to the line $y = x$. Now consider the case of $b = x+it$ merely.

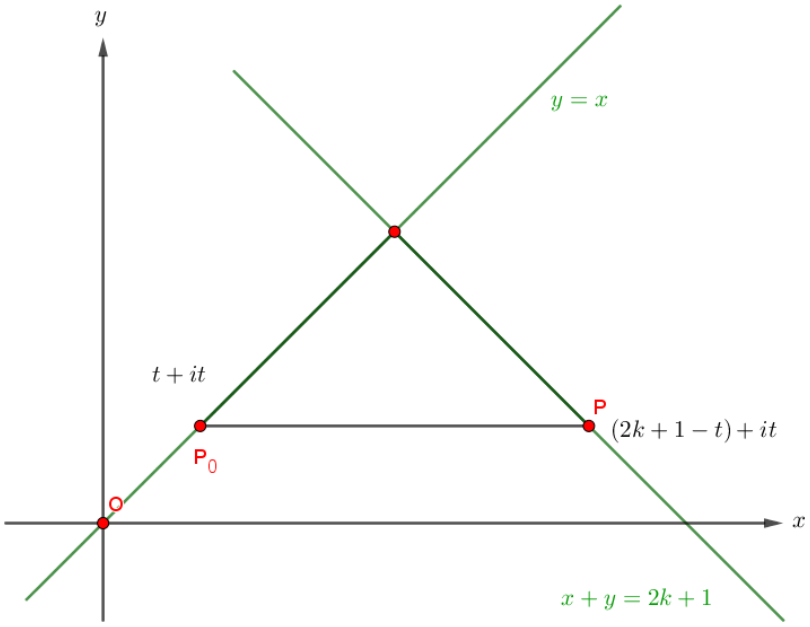


*Figure 1:* Gaussian primes with imaginary part being t

As shown in Figure 1, the lattice $z = (2k+1-t)+it$ is a prime with the imaginary part being $t \geqslant 1$ on the line $l_k : x+y = 2k+1,\ k \in \mathbb{Z}^+$ .Clearly $N(z) \equiv 1 (\mathrm{mod} 4)$ , satisfying the necessary conditions for a Gaussian prime. Without considering the distribution of primes on the coordinate axes, it is obvious that if the norm of an integer is a rational prime of type $4k+1$, then it would be a prime number surely.

Assuming that there are only finitely many such lines $l_k$ , let the farthest such line is $l_{k_M}$ from the lattice point $z_0 = t+it$ , where clearly $k_M > t - \frac{1}{2}$ , then the corresponding







London Journal of Research in Science: Natural and Formal

lattice point with imaginary part being t is $z_M = (2k_M + 1 - t) + it$ , the value of whose norm

$$N(z_M) = (2t^2 - 2t + 1) + 4({k_M}^2 - k_M t + k_M) \tag{3.3}$$

is one rational prime $p > 2$.

Now consider the lattice point $z_M{}'$ with imaginary part being t on the line $l_{k_M+s}$ : $x + y = 2(k_M + s) + 1,\ s \in \mathbb{Z}^+$, whose norm

$$N(z_M{}') = (2t^2 - 2t + 1) + 4[(k_M + s)^2 - (k_M + s)t + (k_M + s))$$

namely,

$$N(z_M{}') = p + 4(s^2 + 2k_M s + s - st) \tag{3.4}$$

where $m = s^2 + 2k_M s + s - st$ is a integer explicitly. By the previous Lemma 1.1, since $(p, 4) = 1$ , to take the appropriate $s > t - (2k_M + 1)$ would make the norm $N(z_M{}') = 4m + p$ a rational prime definitely, and there are infinitely many such ways to do it. This contradicts the hypothesis and the proposition is proved. □

From Proposition 2.4 and the symmetry of the Gaussian primes' distribution, we have the following corollary.

**Corollary 3.2.** *Suppose $t \in \mathbb{Z}$ , there are infinitely many prime numbers on the line $y = t$ or the line $x = t$ .*

### *3.3. Proof of Proposition 2:5.*

*Proof.* By the symmetry of the distribution of primes, we only need to consider the case on the line $l : -x + y = 2k_1 + 1,\ k_1 \in \mathbb{Z}^+$.

The equivalent proposition of Proposition 2.5 is that there exist an infinite number of elementary lines $l_k$ such that their intersection with the line $l$ is Gaussian prime. As shown in Figure 2, obviously, this intersection is $z = (k - k_1) + i(k + k_1 + 1)$ just right, whose norm

$$N(z) = (2k^2 + 2k + 1) + (2{k_1}^2 + 2k_1) \equiv 1(\mathrm{mod}4) \tag{3.5}$$

which satisfies the necessary conditions for one Gaussian prime.

Assuming that there are only finitely many such lines $l_k$ , let the farthest such line is $l_{k_M}$ from the lattice point $z_0 = i(2k_1+1)$ , where clearly $k_M > k_1$ , then the corresponding lattice point on the line $l$ is $z_M = (k_M - k_1) + i(k_M + k_1 + 1)$ , the value of whose norm

$$N(z_M) = (2{k_M}^2 + 2k_M + 1) + (2{k_1}^2 + 2k_1) \tag{3.6}$$

is one rational prime $p > 2$.

Now consider the corresponding lattice point $z_M{}'$ on the line $l_{k_M+s} : x + y = 2(k_M + s) + 1,\ s \in \mathbb{Z}^+$ whose norm




London Journal of Research in Science: Natural and Formal

$$N(z_M{}') = [2(k_M + s)^2 + 2(k_M + s) + 1] + (2{k_1}^2 + 2k_1)$$

that is,

$$N(z_M{}') = p + 2(s^2 + 2k_M s + s_) \tag{3.7}$$

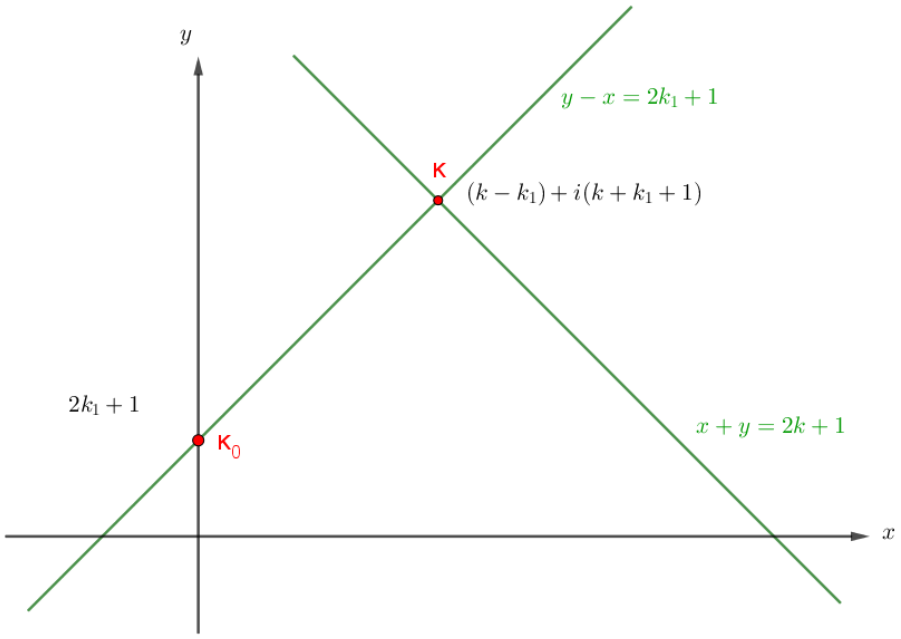


*Figure 2:* Gaussian primes on the line $l : -x + y = 2k_1 + 1,\ k_1 \in \mathbb{Z}^+$

where $m = s^2+2k_M s+s$ is a integer explicitly. By the previous Lemma 1.1, since $(p, 2) = 1$ , to take the appropriate $s > -(2k_M+1)$ would make the norm $N(z_M{}') = 2m+p$ a rational prime definitely. This contradicts the hypothesis and the proposition is proved. □

By the symmetry associated with Corollary 3.1, the same can be obtained for the case on the line $l : x + y = 2k_1 + 1,\ k_1 \in \mathbb{Z}$.

### 3.4. Proof of Proposition 2:6.

*Proof.* Suppose $a(x_1, y_1)$ and $b(x_2, y_2)$ , the line determined by $a, b$ is

$$l' : (y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1) \tag{3.8}$$

Evidently, When the line $l'$ coincides with the line $l_k : x + y = 2k + 1,\ k \in \mathbb{Z}^+$, the conclusion holds according to Proposition 2.5. If they do not overlap, they must have an intersection

$$Q = \left( \frac{x_2(y_2 - y_1) + (2k + 1 - y_2)(x_2 - x_1)}{(x_2 - x_1) + (y_2 - y_1)}, \frac{y_2(x_2 - x_1) + (2k + 1 - x_2)(y_2 - y_1)}{(x_2 - x_1) + (y_2 - y_1)} \right) \tag{3.9}$$







London Journal of Research in Science: Natural and Formal

Clearly, there would be bound to be no infinite number of primes when the line $l$ is one of the lines $l$ ; Now consider the case other than this. It is known that the appropriate $k$ is chosen such that $Q$ is an integer point, and there are an infinite number of such choices. Then Proposition 2.6 can be proved by using the converse method similar to the proof of Propositions 2.4 and 2.5. □

## IV. SEVERAL UNPROVEN PROPOSITIONS

Goldbach's conjecture is formulated as such that every even number not less than 6 is the sum of two odd rational primes.

Let the even number be denoted by $s = 2k(k \geq 3, k \in \mathbb{Z})$ , then the geometric meaning of Goldbach's conjecture is whether, for $\forall k \geq 3$ , there exists a number $0 \leq r < |k|$ on the real number axis such that $p_1 = k + r$ and $p_2 = k - r$ are all the rational primes. That is, whether the midpoint of a line segment joining two the rational primes on the positive x-axis can represent all integer points other than 1 on the positive x-axis. In order to solve such problems, we may wish to develop the following propositions.

**Proposition 4.1.** *All integer points on the complex plane can be represented as the midpoints of two Gaussian primes.*

**Proposition 4.2.** *All integer points on any line $l_k : y = k, k \in \mathbb{Z}^*$ in the complex plane can be represented as the midpoints of two Gaussian primes on that line.*

**Proposition 4.3.** *In the complex plane, all odd points except 1 on the line $l_0 : y = 0$ can be expressed as the midpoints of two Gaussian primes on that line.*

It can be seen that Goldbach's Conjecture would have been proved 'halfway' with the case of that Propositions 4.3 are correct. The fact that the above three propositions are relatively easy to obtain compared to Goldbach's conjecture itself certainly provides a good idea for solving the conjecture.

## V. CONCLUDING REMARKS

This paper explores the distribution of Gaussian primes in the first quadrant from a novel perspective and finally obtains the distribution of Gaussian primes on various types of lines.

As for these last propositions, I have not yet figured out how to prove them.

## REFERENCES


1. Richard A. Mallin. Algebraic number theory [M], CRC Press.
2. Apostol T M . Introduction to Analytic Number Theory[M], Springer-Verlag,1976.
3. I. Kubilyus. The distribution of Gaussian primes in sectors and contours[J]. Leningrad. Gos. Univ. Uc. Zap. Ser. Mat. Nauk 137(19) (1950), 40 {52.